Security and Communication Networks

Machine Learning for Security and Communication Networks 2021

Lead Guest Editor: Chin-Ling Chen
Guest Editors: Lingjuan Lyu

Machine Learning for Security and Communication Networks 2021

Machine Learning for Security and Communication Networks 2021

Lead Guest Editor: Chin-Ling Chen
Guest Editors: Lingjuan Lyu

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9840753, 1 page
https://doi.org/10.1155/2023/9840753

Retraction

Retracted: Detection of DDoS Attack within Industrial IoT Devices Based on Clustering and Graph Structure Features

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Jing and J. Wang, "Detection of DDoS Attack within Industrial IoT Devices Based on Clustering and Graph Structure Features," *Security and Communication Networks*, vol. 2022, Article ID 1401683, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9892175, 1 page
https://doi.org/10.1155/2023/9892175

Retraction

Retracted: The Evaluation of DDoS Attack Effect Based on Neural Network

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Guo, H. Qiu, Z. Liu, J. Zhu, and Q. Wang, "The Evaluation of DDoS Attack Effect Based on Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 5166323, 16 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9820128, 1 page
https://doi.org/10.1155/2023/9820128

Retraction

Retracted: Effective Bots' Detection for Online Smartphone Game Using Multilayer Perceptron Neural Networks

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Tsaur, C. H. Tseng, and C. Chen, "Effective Bots' Detection for Online Smartphone Game Using Multilayer Perceptron Neural Networks," *Security and Communication Networks*, vol. 2022, Article ID 9429475, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9896209, 1 page
https://doi.org/10.1155/2023/9896209

Retraction

Retracted: A Prediction and Evaluation Model Analysis of Enterprise Economic Management Mode Based on Neural Network Strategy

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Wang and X. Han, "A Prediction and Evaluation Model Analysis of Enterprise Economic Management Mode Based on Neural Network Strategy," *Security and Communication Networks*, vol. 2022, Article ID 9639025, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9895458, 1 page
https://doi.org/10.1155/2023/9895458

Retraction

Retracted: Multipurpose Watermarking Approach for Copyright and Integrity of Steganographic Autoencoder Models

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Gu, C. Chang, Y. Bai, Y. Fan, L. Tao, and L. Li, "Multipurpose Watermarking Approach for Copyright and Integrity of Steganographic Autoencoder Models," *Security and Communication Networks*, vol. 2021, Article ID 9936661, 12 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9872735, 1 page
https://doi.org/10.1155/2023/9872735

WILEY | Hindawi

Retraction

Retracted: The Construction of Interactive Classrooms in Colleges and Universities Based on Big Data Analysis and Benchmark Graph Neural Network

Security and Communication Networks

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] P. Shen, "The Construction of Interactive Classrooms in Colleges and Universities Based on Big Data Analysis and Benchmark Graph Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 9214022, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9843539, 1 page
https://doi.org/10.1155/2023/9843539

Retraction

Retracted: Design of Pedagogy Course Information Sharing System Based on Wireless Sensor Network

Security and Communication Networks

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Jin, "Design of Pedagogy Course Information Sharing System Based on Wireless Sensor Network," *Security and Communication Networks*, vol. 2022, Article ID 7040441, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9769165, 1 page
https://doi.org/10.1155/2023/9769165

Retraction

Retracted: An Improved Image Steganography Framework Based on Y Channel Information for Neural Style Transfer

Security and Communication Networks

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Lin, X. Zhu, W. Ye, C. Chang, Y. Liu, and C. Liu, "An Improved Image Steganography Framework Based on Y Channel Information for Neural Style Transfer," *Security and Communication Networks*, vol. 2022, Article ID 2641615, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9843154, 1 page
https://doi.org/10.1155/2023/9843154

Retraction

Retracted: Optimization Model of Logistics Task Allocation Based on Genetic Algorithm

Security and Communication Networks

Received 25 July 2023; Accepted 25 July 2023; Published 26 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Wang and J. Gao, "Optimization Model of Logistics Task Allocation Based on Genetic Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5950876, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9795676, 1 page
https://doi.org/10.1155/2023/9795676

Retraction

Retracted: Empirical Analysis of Enterprise Financial Management Risk Prediction in View of Associative Memory Neural Network

Security and Communication Networks

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Cheng and X. Zhang, "Empirical Analysis of Enterprise Financial Management Risk Prediction in View of Associative Memory Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 7825000, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9768928, 1 page
https://doi.org/10.1155/2023/9768928

Retraction

Retracted: The Application of PSO-SVM Algorithm in the Evaluation System of Sports Competition Events

Security and Communication Networks

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Zhang, "The Application of PSO-SVM Algorithm in the Evaluation System of Sports Competition Events," *Security and Communication Networks*, vol. 2022, Article ID 6865425, 8 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9859568, 1 page
https://doi.org/10.1155/2023/9859568

Retraction

Retracted: Research on Precision Teaching Model of Ideology Course Based on Collaborative Filtering Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Li, "Research on Precision Teaching Model of Ideology Course Based on Collaborative Filtering Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5272779, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9857513, 1 page
https://doi.org/10.1155/2023/9857513

Retraction

Retracted: Research on English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Yi, "Research on English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation," *Security and Communication Networks*, vol. 2022, Article ID 1865451, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9856892, 1 page
https://doi.org/10.1155/2023/9856892

Retraction

Retracted: Research on Prediction of News Public Opinion Guiding Power Based on Neural Network

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Chen and G. Li, "Research on Prediction of News Public Opinion Guiding Power Based on Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 2607492, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9854787, 1 page
https://doi.org/10.1155/2023/9854787

WILEY | Hindawi

Retraction

Retracted: Recognition and Error Correction of Piano Playing Features Based on Filtering Technology

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Fan, "Recognition and Error Correction of Piano Playing Features Based on Filtering Technology," *Security and Communication Networks*, vol. 2022, Article ID 6382743, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9851686, 1 page
https://doi.org/10.1155/2023/9851686

Retraction

Retracted: Machine Learning-Based Emotion Factor Analysis of Sport Fan Community

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Guo, B. Liu, and Z. Yang, "Machine Learning-Based Emotion Factor Analysis of Sport Fan Community," *Security and Communication Networks*, vol. 2022, Article ID 2674987, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9847245, 1 page
https://doi.org/10.1155/2023/9847245

Retraction

Retracted: Machine Learning: The Backbone of Intelligent Trade Credit-Based Systems

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This is article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This is investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] F. Shah, Y. Liu, A. Anwar et al., "Machine Learning: The Backbone of Intelligent Trade Credit-Based Systems," *Security and Communication Networks*, vol. 2022, Article ID 7149902, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9843864, 1 page
https://doi.org/10.1155/2023/9843864

Retraction

Retracted: Research on Enterprise Financial Management and Prediction System Based on SaaS Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Zhang and F. Zhou, "Research on Enterprise Financial Management and Prediction System Based on SaaS Model," *Security and Communication Networks*, vol. 2022, Article ID 3218903, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9842467, 1 page
https://doi.org/10.1155/2023/9842467

Retraction

Retracted: Financial and Economic Sequence Forecasting Based on Time Slot Allocation Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wu, "Financial and Economic Sequence Forecasting Based on Time Slot Allocation Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 2340521, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9841085, 1 page
https://doi.org/10.1155/2023/9841085

Retraction

Retracted: Research Status of Sports Industry Laws from the Perspective of Knowledge Graph

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zhao, "Research Status of Sports Industry Laws from the Perspective of Knowledge Graph," *Security and Communication Networks*, vol. 2022, Article ID 6541921, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9838165, 1 page
https://doi.org/10.1155/2023/9838165

Retraction

Retracted: Analysis of International Competitiveness of China's Mobile Phone Industry Based on Data Mining Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Lu and Z. Lv, "Analysis of International Competitiveness of China's Mobile Phone Industry Based on Data Mining Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 8460574, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9837091, 1 page
https://doi.org/10.1155/2023/9837091

Retraction

Retracted: Multiobjective Optimization Algorithm for EFRM Strategy

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Jiang, X. Wang, and X. Wu, "Multiobjective Optimization Algorithm for EFRM Strategy," *Security and Communication Networks*, vol. 2022, Article ID 4994383, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9826897, 1 page
https://doi.org/10.1155/2023/9826897

Retraction

Retracted: Evaluation System of Foreign Language Teaching Quality Based on Spatiotemporal Feature Fusion

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Hao and Y. Zhu, "Evaluation System of Foreign Language Teaching Quality Based on Spatiotemporal Feature Fusion," *Security and Communication Networks*, vol. 2022, Article ID 9390358, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9824571, 1 page
https://doi.org/10.1155/2023/9824571

Retraction

Retracted: Selective Partial Update Adaptive Filtering Algorithms for Block-Sparse System Identification

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] D. Wei and Q. Xu, "Selective Partial Update Adaptive Filtering Algorithms for Block-Sparse System Identification," *Security and Communication Networks*, vol. 2022, Article ID 2207115, 7 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9823467, 1 page
https://doi.org/10.1155/2023/9823467

Retraction

Retracted: Correlation Analysis of Interbank Money Market Interest Rate and Financial Crisis Based on Neural Network Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Zeng, C. Yuan, and C. Liang, "Correlation Analysis of Interbank Money Market Interest Rate and Financial Crisis Based on Neural Network Model," *Security and Communication Networks*, vol. 2022, Article ID 8807215, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9821647, 1 page
https://doi.org/10.1155/2023/9821647

Retraction

Retracted: The Impact of Financial Development on Agricultural Enterprises in Central China Based on Vector Autoregressive Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Kang and M. Zhao, "The Impact of Financial Development on Agricultural Enterprises in Central China Based on Vector Autoregressive Model," *Security and Communication Networks*, vol. 2022, Article ID 5629202, 16 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9807328, 1 page
https://doi.org/10.1155/2023/9807328

Retraction

Retracted: Emotion Analysis for Foreign Language Learning under Scenarios of Network

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert reader that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zheng and F. Xu, "Emotion Analysis for Foreign Language Learning under Scenarios of Network," *Security and Communication Networks*, vol. 2022, Article ID 2789291, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9806073, 1 page
https://doi.org/10.1155/2023/9806073

Retraction

Retracted: Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Zhang, "Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education," *Security and Communication Networks*, vol. 2022, Article ID 7367418, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9798586, 1 page
https://doi.org/10.1155/2023/9798586

Retraction

Retracted: Exploring the Influence of Big Data Technology on the Innovation of the Enterprise Economic Management Mode

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wu, "Exploring the Influence of Big Data Technology on the Innovation of the Enterprise Economic Management Mode," *Security and Communication Networks*, vol. 2022, Article ID 6241182, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9797213, 1 page
https://doi.org/10.1155/2023/9797213

Retraction

Retracted: Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Long, "Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5418332, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9784859, 1 page
https://doi.org/10.1155/2023/9784859

Retraction

Retracted: Public Sector Performance Assessment Based on DEA Model: The Case of Environmental Policy

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] R. Zhang, "Public Sector Performance Assessment Based on DEA Model: The Case of Environmental Policy," *Security and Communication Networks*, vol. 2022, Article ID 7240379, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9768082, 1 page
https://doi.org/10.1155/2023/9768082

Retraction

Retracted: Cross-Border E-Commerce Logistics Collaboration Model Based on Supply Chain Theory

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Xie, H. Wang, and J. Jiao, "Cross-Border E-Commerce Logistics Collaboration Model Based on Supply Chain Theory," *Security and Communication Networks*, vol. 2022, Article ID 1498765, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9765420, 1 page
https://doi.org/10.1155/2023/9765420

Retraction

Retracted: Application of Digital Image Processing Technology in the Remote Interactive Art Teaching System

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] G. Meng, "Application of Digital Image Processing Technology in the Remote Interactive Art Teaching System," *Security and Communication Networks*, vol. 2022, Article ID 6419117, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9764272, 1 page
https://doi.org/10.1155/2023/9764272

Retraction

Retracted: High-Order Moment Contagion of the Carbon Market: A Heterogeneity Analysis of Market Volatility Trend

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Ni, P. Yun, and J. Sun, "High-Order Moment Contagion of the Carbon Market: A Heterogeneity Analysis of Market Volatility Trend," *Security and Communication Networks*, vol. 2022, Article ID 1608871, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9756418, 1 page
https://doi.org/10.1155/2023/9756418

Retraction

Retracted: Hybrid Music Recommendation Algorithm Based on Music Gene and Improved Knowledge Graph

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Zhang and S. Liu, "Hybrid Music Recommendation Algorithm Based on Music Gene and Improved Knowledge Graph," *Security and Communication Networks*, vol. 2022, Article ID 5889724, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9894212, 1 page
https://doi.org/10.1155/2023/9894212

Retraction

Retracted: Deep-Learning-Based Motion Capture Technology in Film and Television Animation Production

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wei, "Deep-Learning-Based Motion Capture Technology in Film and Television Animation Production," *Security and Communication Networks*, vol. 2022, Article ID 6040371, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9876328, 1 page
https://doi.org/10.1155/2023/9876328

Retraction

Retracted: Construction and Analysis of Urban Cultural Plane Space Mode considering Particle Swarm Cultural Scientific Computing Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Yu and Q. Li, "Construction and Analysis of Urban Cultural Plane Space Mode considering Particle Swarm Cultural Scientific Computing Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 9838782, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9874054, 1 page
https://doi.org/10.1155/2023/9874054

Retraction

Retracted: The Relevance of Microblog Sentiment Analysis on College Students' Growth Development and Management

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Wang, "The Relevance of Microblog Sentiment Analysis on College Students' Growth Development and Management," *Security and Communication Networks*, vol. 2022, Article ID 1797080, 8 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9870482, 1 page
https://doi.org/10.1155/2023/9870482

Retraction

Retracted: Singular Signal Measurement Based on Bidirectional Recursive Complex-Valued Wavelet Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Lv, J. Xie, and J. Huang, "Singular Signal Measurement Based on Bidirectional Recursive Complex-Valued Wavelet Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 9013770, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9870123, 1 page
https://doi.org/10.1155/2023/9870123

Retraction

Retracted: Optimization of Ideological and Political Education Management Strategies under *k*-Means Algorithm in Big Data Environment

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Wang, "Optimization of Ideological and Political Education Management Strategies under *k*-Means Algorithm in Big Data Environment," *Security and Communication Networks*, vol. 2022, Article ID 6120230, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9869657, 1 page
https://doi.org/10.1155/2023/9869657

Retraction

Retracted: A Novel Literary Translation Text Classification Method Based on Distributed Incremental Sequence Data Mining Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Sun, "A Novel Literary Translation Text Classification Method Based on Distributed Incremental Sequence Data Mining Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 1656151, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9869350, 1 page
https://doi.org/10.1155/2023/9869350

Retraction

Retracted: Characteristics Analysis of Applied Mathematics in Colleges and Universities Based on Big Data Mining Algorithm Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wang, F. Tian, and Y. Bai, "Characteristics Analysis of Applied Mathematics in Colleges and Universities Based on Big Data Mining Algorithm Model," *Security and Communication Networks*, vol. 2022, Article ID 7978031, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9865913, 1 page
https://doi.org/10.1155/2023/9865913

Retraction

Retracted: Application of Hidden Markov Model in Financial Time Series Data

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Chang and J. Hu, "Application of Hidden Markov Model in Financial Time Series Data," *Security and Communication Networks*, vol. 2022, Article ID 1465216, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9865853, 1 page
https://doi.org/10.1155/2023/9865853

Retraction

Retracted: Design of College Scheduling Algorithm Based on Improved Genetic Ant Colony Hybrid Optimization

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Li, Q. Xie, and H. Zhang, "Design of College Scheduling Algorithm Based on Improved Genetic Ant Colony Hybrid Optimization," *Security and Communication Networks*, vol. 2022, Article ID 2565639, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9865704, 1 page
https://doi.org/10.1155/2023/9865704

Retraction

Retracted: The Trend Prediction of the New Public Management Model Based on the Discrete Dynamic Evolution Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and name external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Zheng, "The Trend Prediction of the New Public Management Model Based on the Discrete Dynamic Evolution Model," *Security and Communication Networks*, vol. 2022, Article ID 3398392, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9895359, 1 page
https://doi.org/10.1155/2023/9895359

Retraction

Retracted: Digital Protection Technology of Cultural Heritage Based on ArcGIS Geographic Information Technology Algorithm

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Guo, W. Jiang, Q. Zhang, and K. Wang, "Digital Protection Technology of Cultural Heritage Based on ArcGIS Geographic Information Technology Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 3844626, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9876421, 1 page
https://doi.org/10.1155/2023/9876421

Retraction

Retracted: Application of CAD in Semifinished Product Packaging Art Design

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Mao and T. Wu, "Application of CAD in Semifinished Product Packaging Art Design," *Security and Communication Networks*, vol. 2022, Article ID 8937896, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9873758, 1 page
https://doi.org/10.1155/2023/9873758

Retraction

Retracted: Optimization of Printing and Dyeing Energy Consumption Based on Multimedia Machine Learning Algorithm

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Zhang and Y. Yu, "Optimization of Printing and Dyeing Energy Consumption Based on Multimedia Machine Learning Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 1960425, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9862189, 1 page
https://doi.org/10.1155/2023/9862189

Retraction

Retracted: Application of Convolutional Neural Network Algorithm under Deep Learning in Digital Clothing Design

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Xue and X. Xue, "Application of Convolutional Neural Network Algorithm under Deep Learning in Digital Clothing Design," *Security and Communication Networks*, vol. 2022, Article ID 4880555, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9841575, 1 page
https://doi.org/10.1155/2023/9841575

Retraction

Retracted: Art Product Recognition Model Design and Construction of VR Model

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Wu, "Art Product Recognition Model Design and Construction of VR Model," *Security and Communication Networks*, vol. 2022, Article ID 3994102, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9829570, 1 page
https://doi.org/10.1155/2023/9829570

Retraction

Retracted: Application Research of Particle Swarm Algorithm in Bank Human Resource Management

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wu, S. Wang, X. Wang, and Z. Wang, "Application Research of Particle Swarm Algorithm in Bank Human Resource Management," *Security and Communication Networks*, vol. 2022, Article ID 8788894, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9826374, 1 page
https://doi.org/10.1155/2023/9826374

Retraction

Retracted: Practice and Exploration of Teaching Mental Health Education for College Students Based on Data Mining Algorithm

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Bai and C. Tang, "Practice and Exploration of Teaching Mental Health Education for College Students Based on Data Mining Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 8936152, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9818912, 1 page
https://doi.org/10.1155/2023/9818912

Retraction

Retracted: Optimization and Application of Random Forest Algorithm for Applied Mathematics Specialty

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Li, "Optimization and Application of Random Forest Algorithm for Applied Mathematics Specialty," *Security and Communication Networks*, vol. 2022, Article ID 1131994, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9794561, 1 page
https://doi.org/10.1155/2023/9794561

Retraction

Retracted: Risk Assessment Method of Agricultural Management Investment Based on Genetic Neural Network

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zhao, "Risk Assessment Method of Agricultural Management Investment Based on Genetic Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 2373363, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9790486, 1 page
https://doi.org/10.1155/2023/9790486

Retraction

Retracted: Analysis of Efficiency of Human Resource Management Evaluation Model Based on SOM Neural Network

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and name external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Liu, X. Wang, X. Du, and P. Gu, "Analysis of Efficiency of Human Resource Management Evaluation Model Based on SOM Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 4682868, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9781089, 1 page
https://doi.org/10.1155/2023/9781089

Retraction

Retracted: Mental and Emotional Recognition of College Students Based on Brain Signal Features and Data Mining

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. An and Z. Yu, "Mental and Emotional Recognition of College Students Based on Brain Signal Features and Data Mining," *Security and Communication Networks*, vol. 2022, Article ID 4198353, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9853176, 1 page
https://doi.org/10.1155/2023/9853176

Retraction

Retracted: Application of Data Mining Algorithm in Agricultural Products Logistics Network Planning

Security and Communication Networks

Received 30 January 2023; Accepted 30 January 2023; Published 5 February 2023

Security and Communication Networks has retracted the article titled "Application of Data Mining Algorithm in Agricultural Products Logistics Network Planning" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] D.-Y. Yu and G.-L. Wu, "Application of Data Mining Algorithm in Agricultural Products Logistics Network Planning," *Security and Communication Networks*, vol. 2022, Article ID 5884050, 9 pages, 2022.

[2] L. Ferguson, "Advancing research integrity collaboratively and with vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9821518, 1 page
https://doi.org/10.1155/2023/9821518

Retraction

Retracted: Research on Intelligent Recommendation Model of E-Commerce Commodity Based on Feature Selection and Deep Belief Network

Security and Communication Networks

Received 30 January 2023; Accepted 30 January 2023; Published 2 February 2023

Security and Communication Networks has retracted the article titled "Research on Intelligent Recommendation Model of E-Commerce Commodity Based on Feature Selection and Deep Belief Network" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] Y. Li, G. Wu, and C. Liu, "Research on Intelligent Recommendation Model of E-Commerce Commodity Based on Feature Selection and Deep Belief Network," *Security and Communication Networks*, vol. 2022, Article ID 6469217, 11 pages, 2022.
[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9825634, 1 page
https://doi.org/10.1155/2023/9825634

Retraction

Retracted: Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology

Security and Communication Networks

Received 8 December 2022; Accepted 8 December 2022; Published 22 January 2023

Security and Communication Networks has retracted the article titled "Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] Z. Wang and G. Wang, "Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology," *Security and Communication Networks*, vol. 2022, Article ID 4912841, 11 pages, 2022.
[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9805150, 1 page
https://doi.org/10.1155/2022/9805150

Retraction

Retracted: Risk Assessment of Agricultural Economic Management Based on the Multivariate Statistical Computing Method

Security and Communication Networks

Received 12 November 2022; Accepted 12 November 2022; Published 23 November 2022

Security and Communication Networks has retracted the article titled "Risk Assessment of Agricultural Economic Management Based on the Multivariate Statistical Computing Method" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The author does not agree to the retraction.

References

[1] Y. Zhou, "Risk Assessment of Agricultural Economic Management Based on the Multivariate Statistical Computing Method," *Security and Communication Networks*, vol. 2022, Article ID 8547306, 11 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9761517, 1 page
https://doi.org/10.1155/2022/9761517

Retraction

Retracted: A Graph Neural Network (GNN) Algorithm for Constructing the Evolution Process of Rural Settlement Morphology

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 22 November 2022

Security and Communication Networks has retracted the article titled "A Graph Neural Network (GNN) Algorithm for Constructing the Evolution Process of Rural Settlement Morphology" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] Z. Hu, K. Chen, and X. Xie, "A Graph Neural Network (GNN) Algorithm for Constructing the Evolution Process of Rural Settlement Morphology," *Security and Communication Networks*, vol. 2022, Article ID 7517347, 10 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2020, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9795431, 1 page
https://doi.org/10.1155/2022/9795431

Retraction

Retracted: Novel Multirole-Oriented Deep Learning Text Classification Model

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 22 November 2022

Security and Communication Networks has retracted the article titled "Novel Multirole-Oriented Deep Learning Text Classification Model" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] T. Luo, "Novel Multirole-Oriented Deep Learning Text Classification Model," *Security and Communication Networks*, vol. 2022, Article ID 8942841, 11 pages, 2022.
[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9832516, 1 page
https://doi.org/10.1155/2022/9832516

Retraction

Retracted: Feature Extraction and Identification of Calligraphy Style Based on Dual Channel Convolution Network

Security and Communication Networks

Received 10 November 2022; Accepted 10 November 2022; Published 20 November 2022

Security and Communication Networks has retracted the article titled "Feature Extraction and Identification of Calligraphy Style Based on Dual Channel Convolution Network" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] Z. Wang, "Feature Extraction and Identification of Calligraphy Style Based on Dual Channel Convolution Network," *Security and Communication Networks*, vol. 2022, Article ID 4187797, 11 pages, 2022.
[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9791489, 1 page
https://doi.org/10.1155/2022/9791489

Retraction

Retracted: Design and Application of BP Neural Network Optimization Method Based on SIWSPSO Algorithm

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 17 November 2022

Security and Communication Networks has retracted the article titled “Design and Application of BP Neural Network Optimization Method Based on SIWSPSO Algorithm” [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] L. Chu, “Design and Application of BP Neural Network Optimization Method Based on SIWSPSO Algorithm,” *Security and Communication Networks*, vol. 2022, Article ID 2960992, 9 pages, 2022.

[2] L. Ferguson, “Advancing Research Integrity Collaboratively and with Vigour,” 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9845375, 1 page
https://doi.org/10.1155/2022/9845375

Retraction

Retracted: An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 17 November 2022

Security and Communication Networks has retracted the article titled "An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] L. Zhou, Q. Chen, and T. Zhu, "An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations," *Security and Communication Networks*, vol. 2022, Article ID 2173402, 11 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9806073, 1 page
https://doi.org/10.1155/2023/9806073

Retraction

Retracted: Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Zhang, "Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education," *Security and Communication Networks*, vol. 2022, Article ID 7367418, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7367418, 12 pages
https://doi.org/10.1155/2022/7367418

Research Article

Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education

Zehui Zhang

Education Department, Nanchang Normal University, Nanchang 330032, China

Correspondence should be addressed to Zehui Zhang; zhangzehuincnu@seu.edu.mk

Received 7 January 2022; Revised 26 February 2022; Accepted 10 March 2022; Published 23 June 2022

Academic Editor: Chin-Ling Chen

In order to improve the effect of modern music education, this paper applies the digital information technology of music resources to the construction of the music teaching system and derives two new types of semi-decision-making process reinforcement learning algorithms based on the Bellman optimality equation base on discrete time. Moreover, this paper uses the comparative research methods to obtain the Q-value learning curve of the incremental value iterative reinforcement learning algorithm based on the semi-Markov decision process and the incremental value iterative reinforcement learning algorithm based on the dichotomy to improve the fusion effect of music teaching resources. Finally, this paper combines the actual needs of modern music education to construct an intelligent music teaching model.

1. Introduction

Music language plays an important role in the music development of a country. The mother tongue is one or several languages that a person first contacts, learns, and masters [1]. The mother tongue is generally contacted from a young age and continues to be used in adolescents or later. Moreover, in a person's family or formal education, especially in the early stages, a considerable part of the knowledge is imparted through the mother tongue. The language of our country is mainly Chinese, and the music form is mainly Chinese folk songs and the five-tone mode of musical instruments [2]. However, the current universal music language in the world is mainly based on the Western education system, which puts a test for us to understand the world's music culture and creates difficulties for us to convey, learn, and understand the world's excellent music culture. Therefore, modern national educational institutions propose that music culture should inherit and carry forward China's excellent traditional culture, strive to learn the Western music culture knowledge, and integrate the advantages of the two aspects to establish a world cultural form with Chinese cultural heritage. This also puts forward higher requirements for our country's music education in the world and also establishes the world status that our country's music education must achieve [3].

From a broad perspective, the ongoing music education in China is largely based on the level of teachers teaching students to learn, and teachers only understand the teaching of students simply listening to music. This traditional teaching mode has been going through for a long time. This kind of teaching method dominated by the transmission of knowledge is often easy to teach knowledge to death, and it is difficult for students to cultivate character and develop personality. Although I have learned a lot of music knowledge, it is easy to focus on music knowledge and ignore the deeper aspects of music.

At present, the overall development of China's music education industry is facing great challenges, and the trend of international music education cannot be responded quickly; knowledge updating lags behind; music education has not been solved at a higher level; the publication of academic works is extremely rare; China rarely participates in international music education cooperation, and the number of participants is also small. In addition, local traditional music teaching and teacher education in the world, such as oriental music teaching, are considered to be vacant in many fields. It is undeniable that the cultural construction pattern of global integration is inevitable. It has made a positive response to the development of human science and technology, culture, and economy. The living space has been gathered and developed. Therefore, it is not profound for any modern country to abandon its own development or abandon its existing cultural traditions to talk about the development of the world pattern, and it will face many difficulties in the future. Chinese music education first needs to consider the actual background of China's digital music teaching and the development trend of current international music education, such as the mother tongue problem, the problem of education teachers, and the update of knowledge, which have very important historical significance for modern education in my country.

Due to the rapid development of multimedia technology, a large amount of information disseminated on the Internet is in different modalities. For example, for any web page, whether it is about sports, Weibo, or military politics, it will basically contain audio, video, and images, and any one. All modal data information is indispensable. The data information of these different modalities are all developed around the theme of the web page, so cross-modal analysis can be used, that is, the data of different modalities can be found in a certain way to find the correlation between them, so as to carry out comparative analysis. For example, when people listen to a piece of music, the music contains text lyrics that describe the content of the music. These audios and their corresponding text lyrics can be regarded as bimodal information, and they are related to each other in emotional expression. Especially in the era of big data and cloud computing sweeping the network and research in various fields, cross-modal analysis has become a hot spot in this field and has received more and more attention. Therefore, we need to continue to explore and analyze it.

The research and application of traditional single-modal data has been quite extensive, but in the era of multimedia data and big data becoming the mainstream, users' requirements for data are getting higher and higher, and the analysis of single-modal data can no longer keep up with the needs of the times, and people can no longer conduct broader research based on it. Because of the limitation of single modality data, the data information contained in it cannot be fully utilized by people and it cannot describe multimedia data more comprehensively. For multimodal data, they are often complementary to each other, and can express better results when characterizing objects.

The organization structure of this paper is as follows: The first part introduces the related concepts and background of music education and multimodal fusion analysis. The second part is the literature review part, which describes the main work of the music multimodal fusion analysis at home and abroad at present and the music multi-modal fusion analysis. The current situation and difficulties are introduced. The third part proposes an incremental value iterative reinforcement learning algorithm based on the needs of music education, and the algorithm is used to construct a multimodal fusion analysis model for music. The fourth part is based on the support of the third part. The multimodal information fusion music education system is used for learning. The fifth part is to test the validity of the model in this paper by means of experiments. The conclusion part is a summary of the research results of this paper and puts forward the prospect.

The main contributions of this paper are as follows: (1) according to the iterative form of the Bellman optimality equation, a unified analysis framework for SMDP reinforcement learning algorithm is given, which can effectively promote the fusion of multimodal music information; (2) reinforcement learning under the average reward criterion algorithms to carry out related research, so that the music information fusion reinforcement learning algorithm can be more widely used in practical systems.

This paper uses the reinforcement learning model based on information fusion to innovate the way of music education, change the traditional music education model, and improve the actual effect of music education.

2. Related Work

Due to the rapid development of multimedia technology, a large amount of information disseminated on the Internet is of different modalities. For example, for any webpage, whether it is about sports, Weibo, or military politics, it basically contains audio, video, and images, and any one of them. The data information of each modal is indispensable. Moreover, the data information of these different modalities is developed around the theme of the web page, so cross-modal analysis can be used; that is, the data of different modalities can be found in a certain way to find the relationship between them, so as to conduct a comparative analysis. For example, when people listen to a piece of music, the music contains text lyrics that describe the music content. These audio and the corresponding text lyrics can be regarded as bimodal information, and they are related to each other in emotional expression [4]. Especially in the era when big data and cloud computing are sweeping the network and research in various fields, cross-modal analysis has become a hot spot in this field and has attracted more and more attention. Therefore, we need to continue to explore and analyze it. For cross-modal analysis, Cano [5] found the correlation between the two-modal data of text and image through the joint learning of the two-modal data and improved the accuracy of image search, text search, and multi-modal data search. In order to eliminate the duality of semantic expression of text information and improve the accuracy of expression, the text information is correlated with visual characteristics, and the text is expressed more semantically [6]. Dickens [7], based on the Probabilistic Semantic Model (PLSA), learns and trains multimodal information, obtains a hierarchical representation mechanism of multiple modal data, finds the correlation information of data between multiple modalities, and improves data search and query efficiency. Gonçalves [8] based on the multimodal LPP algorithm, trains the music and image sample sets to obtain a common low-dimensional subspace, in which the same sample and different modal data will be close together, that is, multiple modal data. Relevance expression can be carried out, which eliminates the gap in semantic relevance between multiple modal data.

The research and application of traditional single-modal data has been quite extensive, but in the era when multimedia data and big data have become the mainstream, users have higher and higher requirements for data, and the analysis of single-modal data can no longer keep up with the needs of the development of the times. People can no longer carry out a wider range of research based on it. Because of the limitations of single-modal data, the data information it contains cannot be fully utilized by people, and it cannot describe multimedia data more comprehensively. For multimodal data, they are often complementary and interrelated and can express better results when portraying objects [9]. In the research on the automatic generation of multimodal fusion family music albums, by fusing two different modal data of image and music, it played a key role in the semantic understanding of music and images, which has a significant impact on the research results [10]. For multi-modal fusion, Gorbunova [11] compares music and images through the analysis and training of graphs and obtains good results in practical applications. Khulusi [12] based on the principle of Bernoulli distribution aimed to annotate images and texts. Magnusson [13] based on deep learning RBM (Restricted Boltzmann Machines) and DBN (Deep Belief Network) aimed to provide a common feature representation for multimodal data and apply it in various recognition fields. Partesotti [14] used the mutual mapping between text and image to find the mutual relationship between them, and then carried out text identification on the image. Scavone [15] proposed a supervised learning model, which is an improvement on the unsupervised model. It maps multimodal data to a subspace, in which the same type of data will moved closer to each other. In addition, multimodal fusion has also been widely used in classification tasks and the purpose is to improve the accuracy of classification. Serra [16] classifies text, audio, and video features through a fusion model and has achieved good results in the experimental results. Tabuena [17] based on the mapping technology spatially mapped and classified multimodal data, which effectively proves the accuracy of multimodal fusion. Tomašević [18] uses the BOW model to generate file-level lyrics, which are then fused with the extracted audio and video features, and performs semantic classification. In short, multimodal fusion will play an increasingly important role in future research.

The working principle of the emotion model [19] is as follows: first extract emotion words from the emotion library, mainly including some words and sets, and then classify these words and describe the emotions hidden in music according to the classification results. Although this model can better express the semantic information of music, it is not suitable for the extraction of music audio features. Therefore, this model is not suitable for the study of music emotion judgment in this article.

3. Incremental Value Iterative Reinforcement Learning Algorithm Based on Music Education Needs

3.1. Markov Decision Process Reinforcement Learning. Markov decision process reinforcement learning can improve the effect of music information decision-making, so this paper first analyzes the Markov decision process reinforcement learning.

In the discrete-time Bellman optimality equation, since the optimal average return η^{u^*} is unknown, the value iteration algorithm cannot be directly obtained by the formula. Based on the above analysis, it can be seen that the optimal average return η^{u^*} needs to be estimated, and the incremental value iteration algorithm belongs to the direct estimation method. The core idea of the direct estimation method is to directly estimate the optimal average return η^{u^*} in the process of value iteration.

We use $\widetilde{\eta}$ to represent the estimated value of the optimal average return η^{u^*}, and B_{u} and B_{l} represent the upper and lower bounds of the stay $\tau^{u}(i)$. Then, for any strategy $\mu \in \prod_s$ [20],

$$1 \leq B_l \leq n^{\mu} \leq B_u. \tag{1}$$

In the abovementioned formula,

$$n^{\mu} = \pi^{\mu}\tau^{\mu}. \tag{2}$$

It can be seen from the formula that the average return η^{μ} under the semi-Markov decision process can be regarded as the ratio of the average return $\pi^{\mu}r^{\mu}$ and $\pi^{\mu}\tau^{\mu}$. The performance function of the embedded chain is set as follows:

$$f^{\mu} = \frac{1}{n^{\mu}r^{\mu}}. \tag{3}$$

Combining the abovementioned formula, we can get $\pi^{\mu}f^{\mu} = \eta^{\mu}$, so the embedded chain and the original Markov chain have the same average reward. This can be seen as the entire segment is concentrated on the embedded decision point, so the essence of the time-concentrated Markov decision process is to introduce a reward function into the embedded chain. Since the performance function f is not only related to the state but also related to the actions taken in the state, the performance function f cannot be solved directly. In order to be able to perform the iterative process of updating the strategy on the embedded chain, this topic defines the following performance function [21]:

$$f(i,a) = r(i,a) - \widetilde{\eta}\cdot\tau(i,a). \tag{4}$$

The performance potential g^{μ} can be calculated by using the following formula:

$$g^{\mu} = \max_{\mu\in\prod_s}\left\{r^{\mu} - \widetilde{\eta}\tau^{\mu} + P^{\mu}g^{\mu}\right\}. \tag{5}$$

Subsequently, the optimal strategy u* can be obtained using the following strategy update process:

$$\mu^* = \arg\max_{\mu\in\prod_s}\{r^\mu - \tilde{\eta}\tau^\mu + P^\mu g^\mu\}. \quad (6)$$

Through the repeated iteration of formulas (5) and (6), the abovementioned strategy iteration algorithm can be obtained by the optimal strategy under the original semi-Markov decision process. In the iterative process, as the strategy μ approaches the optimal strategy μ^*, the estimated value also continues to approach the optimal average return η^{μ^*}. Therefore, the performance function f(i,a) will also keep changing during the process of strategy iteration.

The optimal strategy ϕ^* of the abovementioned time-focused Markov decision process can be obtained by the following formula [22]:

$$\phi^* = \arg\max_{\mu\in\prod_s}\{\Delta\eta^\mu\}. \quad (7)$$

Among them, there are as follows:

$$\Delta\eta^\mu = \pi^\mu\cdot(r^\mu - \tilde{\eta}\cdot\tau^\mu). \quad (8)$$

$\Delta\eta^\mu$ can be rewritten as follows:

$$\Delta\eta^\mu = (\eta^\mu - \tilde{\eta})\cdot n^\mu. \quad (9)$$

It can be seen from formula (7) that the optimal strategy p is related to the estimated value i of the optimal average return. The optimal strategy is analyzed below.

Since the strategy ui is the optimal strategy under the original semi-Markov decision process, $\eta^{\mu^*} \geq \eta^{\phi^*}$. According to formula (9), we can obtain the following:

$$\Delta\eta^{\phi^*} = \left(\eta^{\phi^*} - \tilde{\eta}\right) \leq \left(\eta^{\mu^*} - \tilde{\eta}\right)\cdot\eta^{\phi^*}. \quad (10)$$

In addition, because the strategy ϕ^* is the optimal strategy of the Markov decision process in time concentration, $\Delta\eta^{\phi^*} \geq \Delta\eta^{\mu^*}$. Using formula (9), we can obtain the following:

$$\Delta\eta^{\phi^*} \geq \left(\eta^{\mu^*} - \tilde{\eta}\right)\cdot\eta^{\mu^*}. \quad (11)$$

Combining formulas (10) and (11), the following relationship can be obtained:

$$\frac{\Delta\eta^{\phi^*}}{n^{\phi^*}} \leq \eta^{\mu^*} - \tilde{\eta} \leq \frac{\Delta\eta^{\phi^*}}{n^{\mu^*}}. \quad (12)$$

If $\tilde{\eta} \geq \eta^{\mu^*}$, the following inequality can be obtained by formula (10):

$$\Delta\eta^{\phi^*} = \left(\eta^{\mu^*} - \tilde{\eta}\right)\cdot n^{\phi^*} \leq 0. \quad (13)$$

According to the abovementioned formula, the following relational formula can be obtained:

$$\Delta\eta^{\phi^*} \leq \tilde{\eta}. \quad (14)$$

If $\tilde{\eta} \leq \eta^{\mu^*}$, combining formulas (9) and (11), we can obtain the following:

$$\Delta\eta^{\phi^*} \geq \eta^{\mu^*} \geq 0. \quad (15)$$

By observing (13) and (15), the following conclusions can be drawn: when $\tilde{\eta} \geq \eta^{\mu^*}$, $\Delta\eta^{\phi^*} \leq 0$ is established, when $\tilde{\eta} \leq \eta^{\mu^*}$, $\Delta\eta^{\phi^*} \geq 0$ is established. In summary, $\Delta\eta^{\phi^*}$ can provide the correct search direction for the optimal average return. Subsequently, the estimated value η^{μ^*} of the optimal average return $\tilde{\eta}$ can be updated iteratively through one-dimensional search, and the process is as follows:

$$\tilde{\eta} \leftarrow \tilde{\eta} + H\cdot\Delta\eta^{\phi^*}, H = \frac{1}{B_\mu}. \quad (16)$$

In the abovementioned formula, H is the search step size of the estimated value $\tilde{\eta}$.

3.2. SMDP Incremental Iterative Reinforcement Learning Algorithm. The extended form of the SMDP incremental value iterative reinforcement learning algorithm is the same as the R-learning algorithm, which is derived from the discrete-time Bellman optimality equation. Therefore, the SMDP incremental value iterative reinforcement learning algorithm's state-action-value function update formula is the same as that of the R-learning algorithm, and the specific process is shown in Algorithm 1.

(1) Initialize Q (state action to Q-value) expressed as any real number, usually make Q0(i, a) = 0, $\forall(i,a) \in S\times A$.

The algorithm sets $t_m = t_n = 0$ and specifies $\varepsilon > \sigma > 0$ and initializes $\tilde{\eta}_0$ as any real number.

(2) The algorithm calculates $Q_{t_m+1}(X_n, A_n)$ for each state $X_n \in S$:

$$Q_{t_m+1}(X_n, A_n) \leftarrow (1-\alpha)\cdot Q_{t_m}(X_n, A_n) + \alpha\cdot\left(R(X_n, A_n) - \tilde{\eta}\cdot\tau(X_n, A_n) + \max_{a'\in A} Q_{t_m}(X_{n+1}, a')\right). \quad (17)$$

(3) If $sp(Q_{t_m+1} - Q_{t_m}) \leq \sigma$, the algorithm executes step 4. *sp(Q)* is the span of Q, as shown below:

$$sp(Q) \equiv \underset{i\hat{I}S}{max}\,\underset{a\hat{I}A(i)}{max}\,Q(i,a) - \underset{i\hat{I}S}{min}\,\underset{a\hat{I}A(i)}{max}\,Q(i,a). \quad (18)$$

Otherwise, the algorithm sets $t_m \leftarrow t_m + 1$, $n \leftarrow n+1$ and executes Step 2.

(4) The algorithm calculates $\Delta\eta^{\phi^*}$:

$$\frac{1}{2}\cdot\left[\max_{i\in S}\left(\max_{a\in A(i)} Q_{t_m+1}(i,a)-\max_{a\in A(i)} Q_{t_m}(i,a)\right)\right.$$
$$\left.+\min_{i\in S}\left(\max_{a\in A(i)} Q_{t_m+1}(i,a)-\max_{a\in A(i)} Q_{t_m}(i,a)\right)\right]. \tag{19}$$

If $|\Delta\eta^{\phi^*}|\leq\varepsilon$, the algorithm executes Step 5. Otherwise, the algorithm updates $\tilde{\eta}$.

$$\tilde{\eta}_{t_m+1}\leftarrow\tilde{\eta}_{t_m}H\cdot\Delta\eta^{\phi^*}, H=\frac{1}{B_\mu}. \tag{20}$$

The algorithm sets up $t_n\leftarrow t_n+1$, $t_m\leftarrow t_m+1$, $n\leftarrow n+1$ and returns to Step 2.

(5) For any state i s $\in$ S, the algorithm selects the optimal strategy μ^* through the following formula:

$$\mu^*(i)\in\arg\max_{a\in A(i)} Q_{t_m}(i,a). \tag{21}$$

Inspired by formula (16), since $\Delta\eta^{\phi^*}$ can provide the correct search direction for the optimal average return, we can use the dichotomy to replace formula (16) to estimate the optimal average return directly. The algorithm gives two initial estimates element and element, so that it satisfies the relationship $\tilde{\eta}_1\leq\eta^{\mu^*}\leq\tilde{\eta}_2$. The algorithm uses $\tilde{\eta}\leq\eta^{\mu^*}\leq\tilde{\eta}_2=(\tilde{\eta}_1+\tilde{\eta}_2)$ as the initial estimate of the optimal average return $\Delta\eta^{\mu^*}$. If $\Delta\eta^{\phi^*}>0$, the algorithm sets $\tilde{\eta}_1\leftarrow\tilde{\eta}$; otherwise, the algorithm sets $\tilde{\eta}_2\leftarrow\tilde{\eta}$. Using the abovementioned coin update method instead of formula (20) in algorithm 1, an iterative reinforcement learning algorithm based on the dichotomy of incremental values can be obtained. Due to the introduction of the dichotomy, the IVI reinforcement learning algorithm based on the dichotomy also has a higher optimization efficiency in the case of a large state action space. Algorithm flowchart is shown in Figure 1.

3.3. Reinforcement Learning Curve Analysis. This section gives the Q-value learning curve of the incremental value iterative reinforcement learning algorithm for the semi-Markov decision process and the incremental value iterative reinforcement learning algorithm based on the dichotomy. The simulation experiment results of these two algorithms are shown in Figures 1 and 2, respectively. The learning rate α of the incremental value iterative reinforcement learning algorithm and the incremental value iterative reinforcement learning algorithm based on dichotomy are both $1/(k(X_n,A_n)+1)$, and the parameters s and α are 10^* and 10, respectively. The search step size of the incremental value iterative reinforcement learning algorithm is $H=0.1$, and the initial optimal average return estimate $\tilde{\eta}_0=4$. The initial estimates of the incremental value iterative reinforcement learning algorithm based on dichotomy $\tilde{\eta}_1$ and $\tilde{\eta}_2$ are $\tilde{\eta}_1=0$ and $\tilde{\eta}_2=6$, respectively, and the initial estimated value of the optimal average return η^{μ^*} is $\tilde{\eta}=3$.

As can be seen from Figures 2(a) and 2(b), the IVI reinforcement learning algorithm and the dichotomy-based IVI reinforcement learning algorithm both converge within 40,000 iteration steps. Among them, the IVI reinforcement learning algorithm converges faster than the IVI reinforcement learning algorithm based on the dichotomy.

In this paper, the incremental value iteration algorithm is used to directly estimate the optimal average return. In addition, this section uses the SSP value iterative algorithm to directly estimate the optimal average reward, so as to obtain the SMDP random shortest path value iterative reinforcement learning algorithm. The specific process of the random shortest path problem is shown in Figure 3.

As can be seen from the abovementioned figure, the random shortest path problem sets the transition probability $p^{\mu(i)}(i,\xi)$ to a special state ξ to zero, while the transition probability $p^{\mu(i)}(i,\xi)$ in other states remains unchanged. Artificially introduce the absorption termination state t, and set the transition probability of any state i to the termination state t as $p^{\mu(i)}(i,\xi)$. The expected reward function $r^{\mu}_{SSP}(i)$ of the random shortest path problem under the state $i\in S$ is defined as follows:

$$r^{\mu}_{SSP}(i)=r^{\mu}(i)-\tilde{\eta}\cdot\tau^{\mu}(i). \tag{22}$$

The algorithm assumes that $g_\mu(i)$ represents the expected total reward of the random shortest path problem under strategy u starting from state i. Subsequently, there is the following relationship:

$$g_\mu(i)-\sum_{j=1}^{\xi-1}p^{\mu}(i,j)\cdot g_\mu(i)=r^{\mu}(i,\mu(i))-\tilde{\eta}\cdot\tau^{\mu}(i,\mu(i)). \tag{23}$$

Both sides of the abovementioned formula are multiplied by $\pi^{\mu}(i)$ to the left at the same time, and then they are accumulated and summed from $i=1$ to ξ to obtain the following formula:

$$\pi^{\mu}(i)\cdot g_\mu(i)=\pi^{\mu}(i)\cdot\left[r^{\mu}(i,\mu(i))-\tilde{\eta}\cdot\tau^{\mu}(i,\mu(i))\right]=\Delta\eta^{\mu}. \tag{24}$$

From the abovementioned formula, the following relationship can be derived:

$$g_\mu(\xi)=\frac{\Delta\eta^{\mu}}{\pi^{\mu}(\xi)}. \tag{25}$$

The optimal strategy φ^* for the random shortest path problem can be obtained by the following formula:

$$\varphi^*=\arg\max_{\mu\in\prod_s} g_\mu(\xi). \tag{26}$$

The derivation process of the iterative update formula of the SSP value iterative algorithm coin is similar to the algorithm. Combining (2) and (9), we can see the following:

$$\Delta\eta^{\mu}=\pi^{\mu}\tau^{\mu}\cdot(\eta^{\mu}-\tilde{\eta}). \tag{27}$$

According to the optimality of the strategy u*, we know that $\eta^{\mu^*}\geq\eta^{\varphi^*}$. Combining formulas (25) and (27), we can obtain the following:

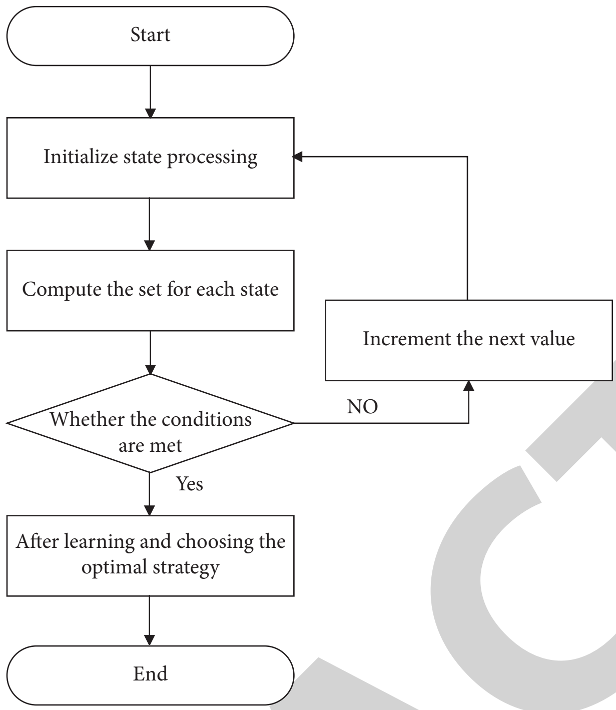
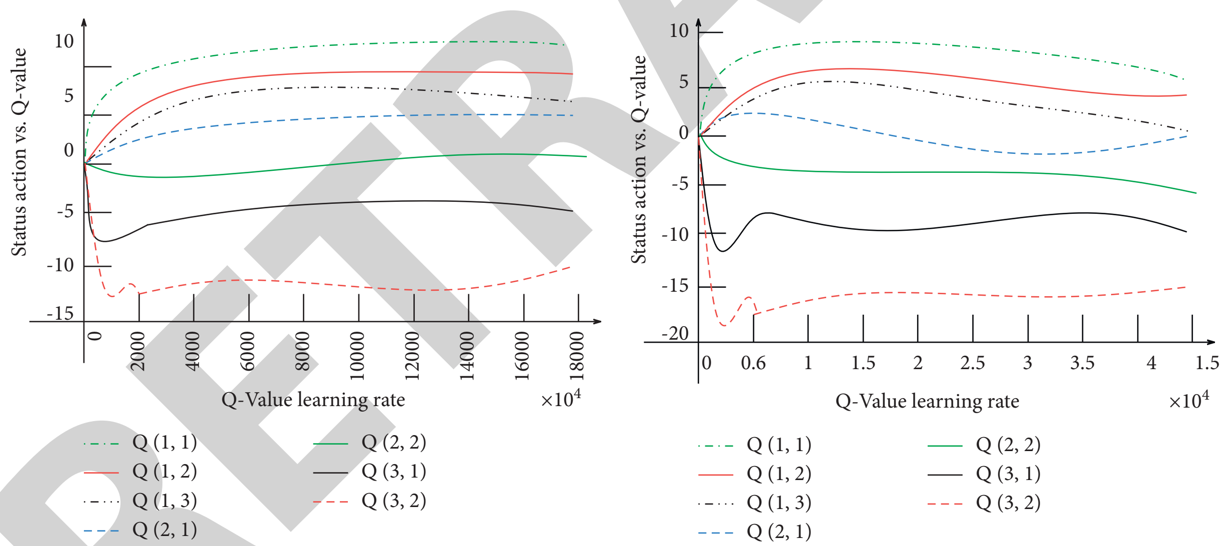


Figure 1: Algorithm flowchart.



(a)

(b)

Figure 2: Learning curve. (a) Q-value learning curve of IVI reinforcement learning algorithm. (b) Q-value learning curve of IVI reinforcement learning algorithm based on dichotomy.

$$g_{\varphi^*}(\xi) = \frac{\Delta\eta^{\varphi^*}}{\pi^{\varphi^*}(\xi)} = \frac{\pi^{\varphi^*}\tau^{\varphi^*}}{\pi^{\varphi^*}(\xi)}\cdot\left(\eta^{\varphi^*}-\tilde{\eta}\right) \le \frac{\pi^{\varphi^*}\tau^{\varphi^*}}{\pi^{\varphi^*}(\xi)}\cdot\left(\eta^{\varphi^*}-\tilde{\eta}\right). \quad (28)$$

According to the optimality of strategy φ^*, we know that $\Delta\eta^{\varphi^*} \ge \Delta\eta^{\mu^*}$. Similar to the above analysis, we can obtain the following:

$$g_{\varphi^*}(\xi) = \frac{\Delta\eta^{\varphi^*}}{\pi^{\varphi^*}(\xi)} \ge \frac{\Delta\eta^{\mu^*}}{\pi^{\varphi^*}(\xi)} = \frac{\pi^{\varphi^*}\tau^{\varphi^*}}{\pi^{\varphi^*}(\xi)}\cdot\left(\eta^{\mu^*}-\tilde{\eta}\right). \quad (29)$$

Under the assumption of ergodicity, according to (28) and (29), it can be known that when $\tilde{\eta} \ge \eta^{\mu^*}$, $g_{\varphi^*}(\xi) \le 0$ holds, and when $\tilde{\eta} \le \eta^{\mu^*}$, $g_{\varphi^*}(\xi) \ge 0$ holds. In summary, $g_{\varphi^*}(\xi)$ can provide the correct search direction for the optimal average return η^{μ^*}. Therefore, the estimated value $\tilde{\eta}$ of η^{μ^*} can be updated iteratively by the following formula:

$$\tilde{\eta} \leftarrow \tilde{\eta} + \gamma\cdot g_{\varphi^*}(\xi), \gamma = \frac{1}{B_\mu}. \quad (30)$$

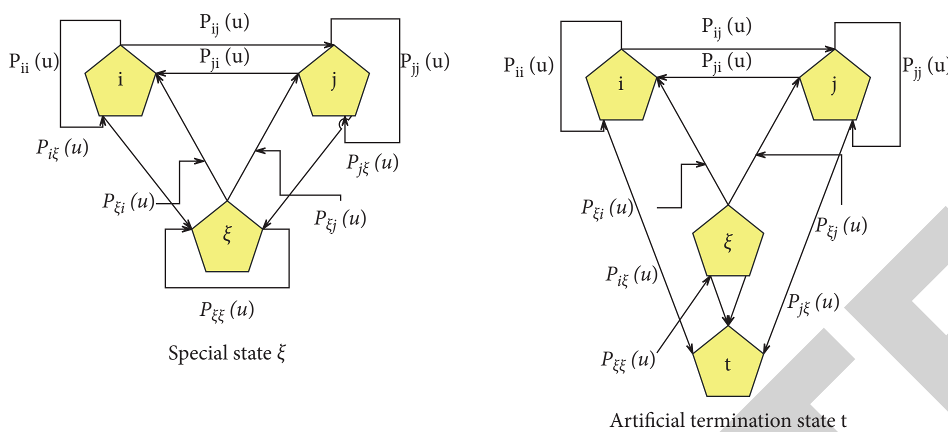


Figure 3: Example of a random shortest path problem.

In the abovementioned formula, γ is the search step length of the estimated value $\tilde{\eta}$. Subsequently, this subject obtained the random shortest path value iteration reinforcement learning algorithm under the semi-Markov decision process.

(1) The algorithm initializes the Q table (state action versus Q-value) to any real number, and the algorithm usually sets $Q_0(\mathrm{i,a}) = 0,\ \forall (i, a) \in S \times A$.

The algorithm sets $t_{\mathrm{m}} = t_{\mathrm{n}} = 0$ and specifies $\varepsilon > \sigma > 0$ and initializes $\tilde{\eta}_0$ as any real number.

(2) For each state $X_n \in S$, the algorithm calculates $Q_{t_m+1}(X_n, A_n)$:

$$Q_{t_m+1}(X_n, A_n) \leftarrow (1-\alpha)\cdot Q_{t_m}(X_n, A_n) + \alpha\cdot\left(R(X_n, A_n) - \tilde{\eta}\cdot\tau(X_n, A_n) + \max_{a'\in A} Q_{t_m}(X_{n+1}, a')\cdot I\{X_{n+1}\neq\xi\}\right). \tag{31}$$

Among them, $I\{\ldots\}$ is the indicator function.

(3) If $sp(Q_{t_m+1} - Q_{t_m}) \leq \sigma$, the algorithm executes step 4. Otherwise, the algorithm sets $t_m \leftarrow t_m + 1$, $n \leftarrow n + 1$ and executes step 2.

(4) The algorithm uses $\max_{a\in A(i)} Q_{t_m+1}(\xi, a)$ to calculate $g_{\varphi^*}(\xi)$. If $|\max_{a\in A(i)} Q_{t_m+1}(\xi, a)| \leq \varepsilon$, the algorithm executes step 5. Otherwise, the algorithm updates $\tilde{\eta}$.

$$\tilde{\eta}_{t_n+1} \leftarrow \tilde{\eta}_{t_n} + \gamma\cdot\max_{a\in A} Q_{t_m+1},\ \gamma = \frac{1}{B_\mu}. \tag{32}$$

We set $t_n \leftarrow t_n + 1$, $t_m \leftarrow t_m + 1$, $n \leftarrow n + 1$ and return to Step 2.

(5) For any state $i \in \mathrm{S}$, the algorithm selects the optimal strategy μ^* by the following formula.

$$\mu^*(i) \in \arg\max_{a\in A(i)} Q_{t_m}(i, a). \tag{33}$$

Since $g_{\varphi^*}(\xi)$ can provide the correct search direction for the optimal average reward η^{μ^*}, the dichotomy can also be used to replace formula (30) to directly estimate the optimal average reward η^{μ^*}. The algorithm gives initial estimates $\tilde{\eta}_1$ and $\tilde{\eta}_2$, if $g_{\varphi^*}(\xi) > 0$, the algorithm sets $\tilde{\eta}_1 \leftarrow \tilde{\eta}$, otherwise, the algorithm sets $\tilde{\eta}_2 \leftarrow \tilde{\eta}$. Using the abovementioned update method instead of formula (32) in algorithm 2, an iterative reinforcement learning algorithm based on the dichotomy of SSP values can be obtained.

The Q-value learning curve of the SMDP random shortest path value iterative reinforcement learning algorithm and the random shortest path value iterative reinforcement learning algorithm based on the dichotomy are given, respectively. The simulation experiment results of the algorithm are shown in Figures 4 and 5, respectively. The simulation environment of these two algorithms is the same as that of the IVI reinforcement learning algorithm. The learning rate of SSP value iterative reinforcement learning algorithm α, parameters s and o, and the estimated value of the initial optimal average return are the same as the IVI reinforcement learning algorithm, and its search step $\gamma = 0.1$. The initial estimates $\tilde{\eta}_1$ and $\tilde{\eta}_2$ of the SSP value iterative reinforcement learning algorithm based on the dichotomy are the same as those of the IVI reinforcement learning algorithm based on the dichotomy. The special state in the random shortest path problem is $\xi = 1$.

It can be seen from Figures 4 and 5 that the SSP value iterative reinforcement learning algorithm and the dichotomy-based SSP value iterative reinforcement learning algorithm both converged within 22,000 iteration steps. Among them, the SSP value iterative reinforcement learning algorithm converges faster than the SSP value iterative reinforcement learning algorithm based on the dichotomy.

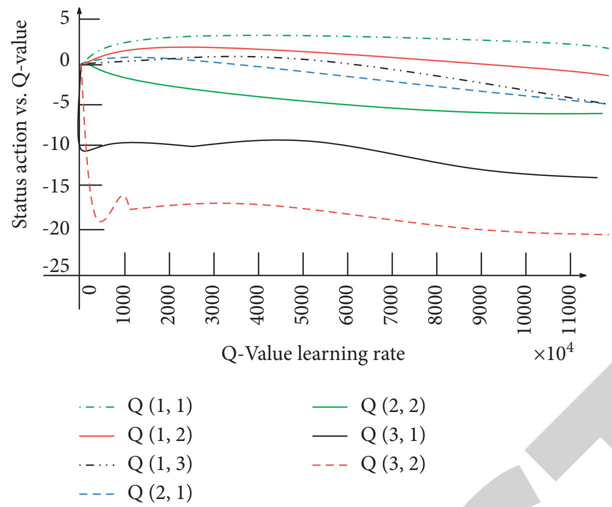


Figure 4: Q-value learning curve of SSP value iterative reinforcement learning algorithm.

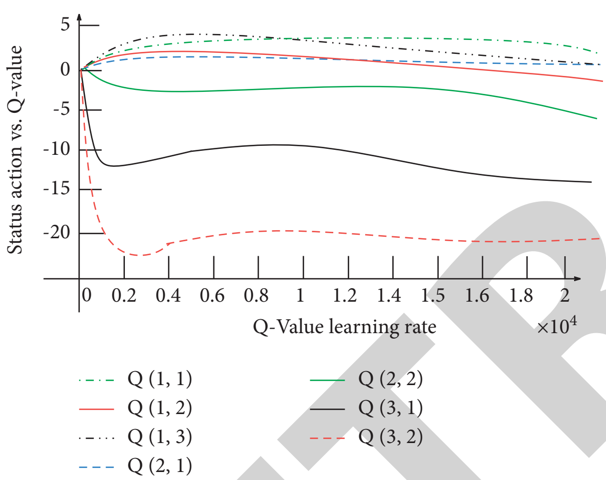


Figure 5: Q-value learning curve of SSP value iterative reinforcement learning algorithm based on dichotomy.

The SSP value iterative reinforcement learning algorithm has the fastest convergence speed and the best convergence performance. Compared with other semi-Markov reinforcement learning algorithms in this article, the SSP value iterative reinforcement learning algorithm has the highest efficiency in finding the optimal strategy. By comparing Figures 1 and 4, it can be seen that the SSP value iterative reinforcement learning algorithm has improved the convergence performance of the IVI reinforcement learning algorithm. The simulation results in Figures 4 and 5 verify the convergence of the SSP value iterative reinforcement learning algorithm and the SSP value iterative reinforcement learning algorithm based on the dichotomy. Although the effect of using the dichotomy to directly estimate the optimal average return η^{μ^*} is not as good as the SSP value iterative reinforcement learning algorithm. However, the iterative reinforcement learning algorithm of SSP value based on dichotomy compared with the semi-Markov reinforcement learning algorithm in this paper still has a large degree of improvement in convergence performance.

4. Music Education Innovation Model Based on Information Fusion Reinforcement Learning Model

The multimedia system for studying music teaching needs to meet the following basic requirements. First, it needs to build a rich music resource library, which is the core of the multimedia teaching system. Second, for students with a higher start, the system needs to provide high-quality music works for students to appreciate and learn, which will help improve students' music appreciation level. For students with a low start, the system needs to provide a variety of basic teaching resources to lay a solid foundation for students.

There are video resources, document resources, and picture resources in the music multimedia teaching system. The system needs to support online browsing of different types of music resources. In addition, it needs to provide ways to realize the unified browsing of video resources, document resources, and picture resources.

In addition, the system needs to support intelligent answering. Traditional teaching systems only support real-time interaction between teachers and students. This interaction mode is limited by time and space, and teachers cannot answer students' questions if they are not online. In order to solve this defect, the system introduces intelligent Q&A, which can solve most students' questions and help improve the quality of teaching.

Music resources must have a detailed classification system, which is helpful for subsequent location searches and allows students to quickly find music resources of interest.

Once the music resources stored in the traditional paper mode are damaged, they cannot be restored. However, the resources stored after informatization must have high reliability and be able to deal with database failures under normal circumstances. The basic requirements can be summarized from the above requirements, such as dynamic management of music resources and music resource

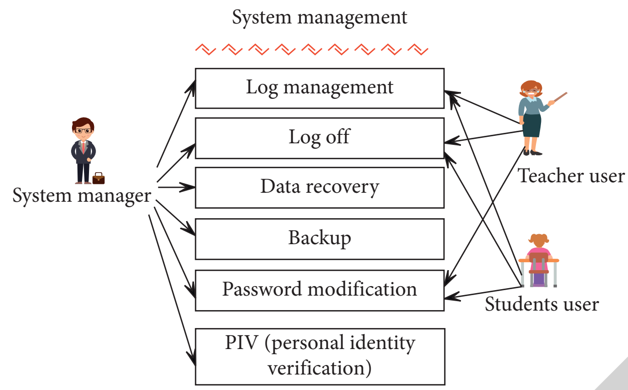


Figure 6: Use case diagram of system management module.

classification. At the same time, data backup and data recovery functions need to be introduced to save music resources. The abovementioned analysis is the basis for clarifying the functional requirements of the music multimedia teaching system.

The use case diagram of the system management module is shown in Figure 6.

The user can accurately locate the music resources in the system according to the resource keywords to save time. If the user searches for the resources one by one, a lot of time will be wasted. The use case diagram of the music resource management module is shown in Figure 7.

Homework management is mainly to examine students' mastery of each knowledge point. Teachers can post word-based homework online, and students can upload their answers online to the teacher for correction after completion. The use case diagram of the homework management module is shown in Figure 8.

The purpose of building a music multimedia teaching system in this subject is to realize the sharing of music resources. The physical structure of the system is shown in Figure 9.

The music multimedia teaching system is designed and implemented based on the J2EE platform. Developers can assemble commonly used codes in the system to form reusable codes so that they can be reused. The architecture diagram is shown in Figure 10.

The music multimedia teaching system is composed of five parts: music resource management, system management, homework management, interactive management, and basic data setting. The functional structure diagram is shown in Figure 11.

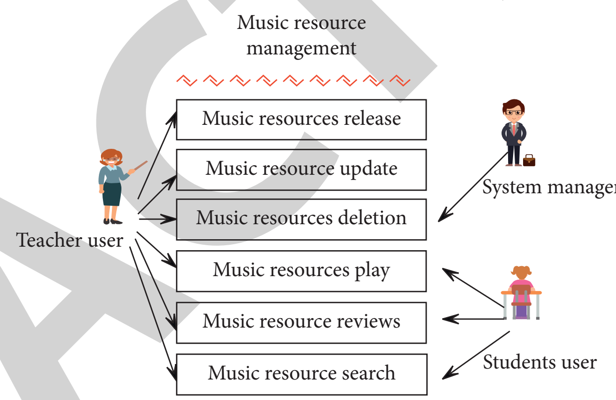


Figure 7: Use case diagram of music resource management module.

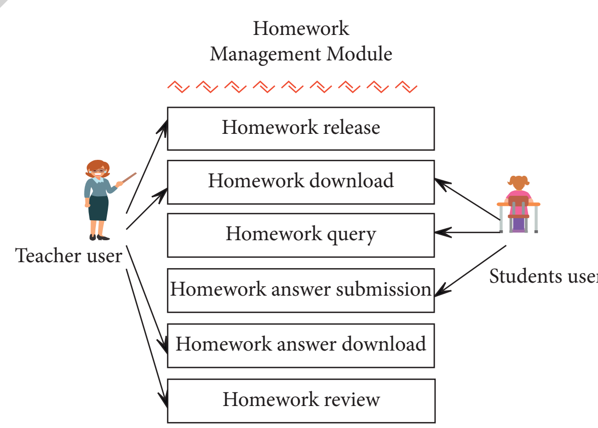


Figure 8: Use case diagram of homework management module.

5. Test and Discussion

MATLAB-based expansion test analysis: the specific implementation process of the experiment is to lock the source of the data on the music platform and collect it based on the target of the task.

On the basis of the abovementioned analysis, the model proposed in this paper is verified, and the practical effects of the music education innovation system based on the information fusion reinforcement learning model are explored, and the information fusion and teaching effects of the system are, respectively, counted. The method proposed in this paper is compared with the literature in [20]. The results are shown in Tables 1 and 2.

From the abovementioned research, it can be seen that the music education innovation system based on the information fusion reinforcement learning model proposed in this paper has a good music resource information fusion effect and teaching effect.

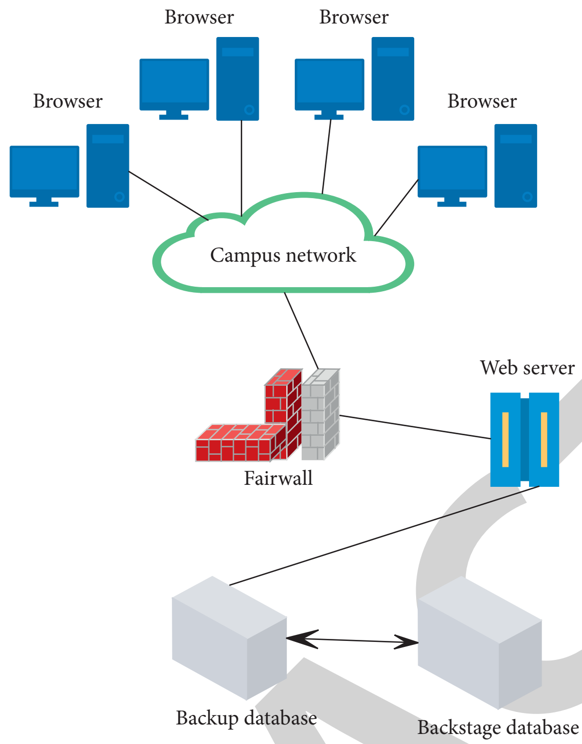


Figure 9: System topology diagram.

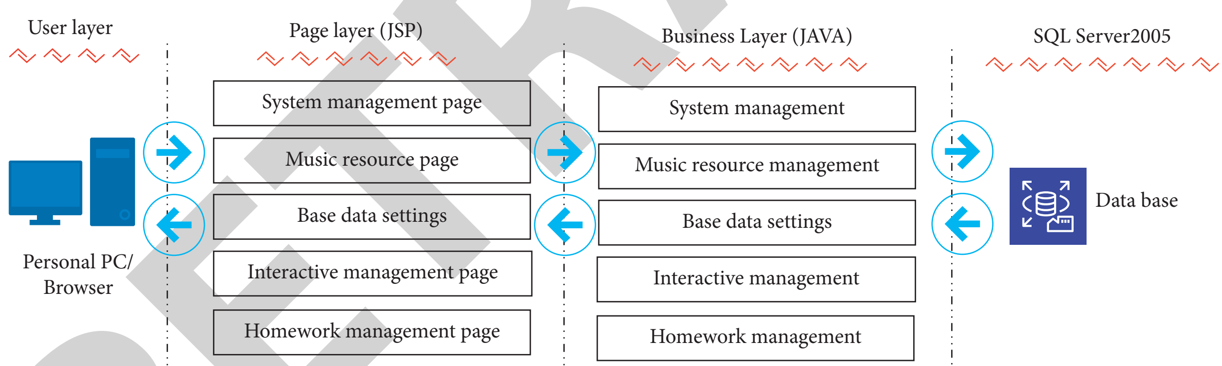


Figure 10: Structure of the system.

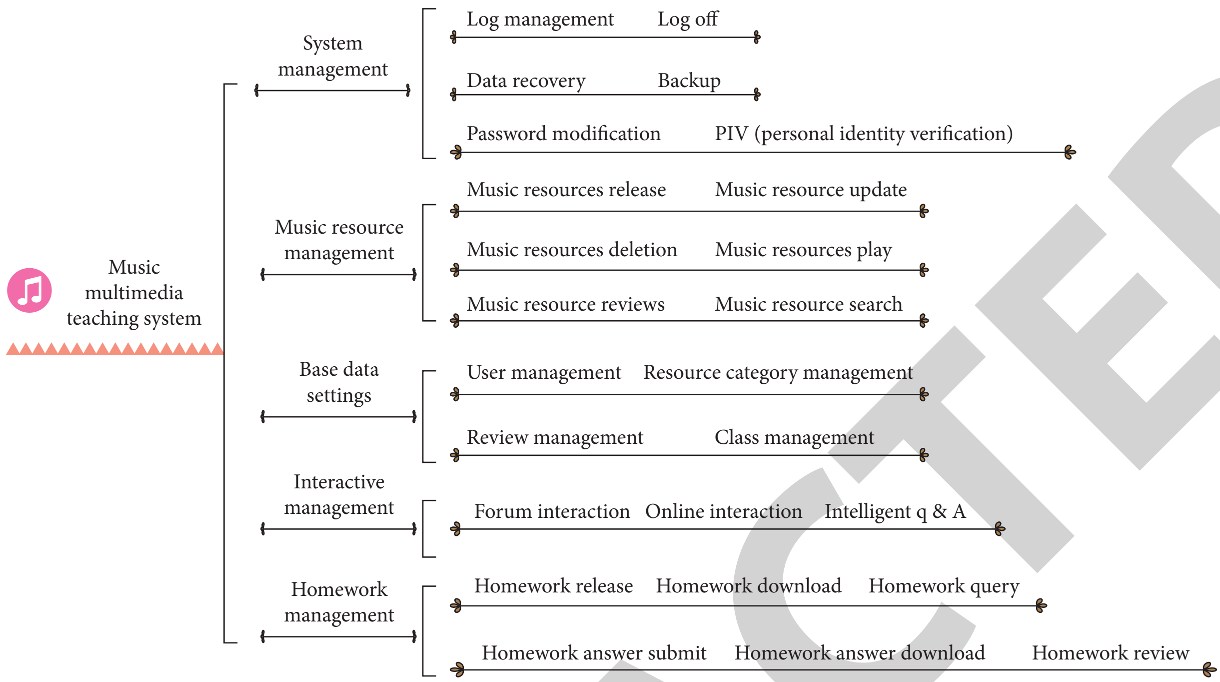


Figure 11: Functional structure diagram.

Table 1: Evaluation of the effect of music resource information fusion.

Number	The method of this paper	The method of [20]	Number	The method of this paper	The method of [20]
1	91.47	81.91	16	95.70	95.27
2	93.68	81.13	17	91.52	83.11
3	94.46	81.63	18	89.09	86.11
4	89.54	79.00	19	96.91	83.27
5	93.84	84.77	20	89.64	88.15
6	92.41	89.93	21	91.25	89.53
7	92.19	80.55	22	92.31	89.91
8	91.98	87.33	23	91.21	84.25
9	92.72	90.38	24	96.39	95.97
10	88.03	87.57	25	93.63	86.97
11	94.82	84.14	26	92.42	83.82
12	92.94	80.34	27	92.29	87.45
13	96.46	83.99	28	95.40	85.42
14	92.62	86.47	29	91.22	90.57
15	91.15	89.30	30	92.93	84.80

Table 2: Evaluation of the effect of music education innovation.

Number	The method of this paper	The method of [20]	Number	The method of this paper	The method of [20]
1	77.74	82.93	16	71.86	82.58
2	89.49	92.67	17	79.80	80.54
3	77.50	82.66	18	91.72	78.29
4	85.37	89.41	19	85.62	90.26
5	72.11	84.10	20	87.92	88.83
6	80.13	86.20	21	76.53	84.52
7	73.44	85.58	22	81.72	88.32
8	82.07	86.25	23	85.94	88.87
9	85.76	79.19	24	90.45	84.80
10	72.06	80.71	25	76.27	84.42
11	88.56	91.30	26	87.21	89.19
12	90.34	85.12	27	82.15	86.11
13	79.23	92.26	28	81.14	88.42
14	72.33	85.06	29	71.37	80.05
15	85.70	85.46	30	71.96	79.25

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9841085, 1 page
https://doi.org/10.1155/2023/9841085

Retraction

Retracted: Research Status of Sports Industry Laws from the Perspective of Knowledge Graph

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zhao, "Research Status of Sports Industry Laws from the Perspective of Knowledge Graph," *Security and Communication Networks*, vol. 2022, Article ID 6541921, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6541921, 11 pages
https://doi.org/10.1155/2022/6541921

Research Article

Research Status of Sports Industry Laws from the Perspective of Knowledge Graph

Jialei Zhao

Department of Physical Education, Taiyuan Normal University, Taiyuan, Shanxi, China

Correspondence should be addressed to Jialei Zhao; zhaojialei2022@tynu.edu.cn

Received 13 April 2022; Revised 24 May 2022; Accepted 7 June 2022; Published 23 June 2022

Academic Editor: Chin-Ling Chen

The purpose of this study is to explore the hot topics of sports industry policy research and the future trends of research in this field. The time node selects the literature published in related journals from 1990 to 2022 as sample data. The research uses citespace6.1 software to draw a knowledge map for research in this field, relevant statistics of research literature in this field, mathematical statistics, word frequency analysis, keyword cooccurrence analysis, and other methods. This study analyzed the temporal and spatial distribution of literature, the specific situation of cooperative research, the changes of research hotspots, and the future research trend in the whole development process of sports industry policy research. Results show that sports colleges/institution is the main institution for sports industry policy research. Among them, sports colleges and professional sports colleges in comprehensive universities occupy a dominant position in research results; research strength is mainly concentrated in the eastern and central regions, and the core authors of the article are mainly concentrated in the Middle East; the cooperation situation of sports industry policy research is gradually formed, but the scale and degree of cooperation in this field have a large room for improvement at present; the main research hotspots in this field are sports industry, sports industry policy, sports consumption, etc.; the hotspots of sports industry policy research are differentiated according to different time periods.

1. Introduction

As an important part of the national economy, the sports industry plays a major role in adjusting the economic structure, driving economic growth, expanding employment channels, and promoting the development of related industries. Policies are indispensable support for the development of the sports industry. Relevant policies play a vital role in the development direction of the sports industry and the grasp of relevant development priorities and hot spots. This research will be of great significance to the literature on sports industry policy research in China and analyze the changes in research hotspots in sports industry policy. Research trends in the field and clarifies the development context for related research on China's sports industry policy.

The sports industry policy is one of the national economic policies. It is formulated by the national or regional government to achieve the goal of social development. It acts as an economic lever to intervene and regulate the development of the sports industry [1, 2]. China's sports industry policy appeared after the emergence and development of the sports industry. The earliest literature on sports industry policy in China appeared in 1994. China's sports industry policy-related research has been increasing year by year since 2000. Since 2010, sports industry policy research has received more attention, and in 2016, the research fever reached its peak [3, 4]. Despite this, there is relatively little content in this stage of research on how to formulate industrial policies, how policies are transmitted, how they are controlled, how to supervise the implementation of policies, and the performance evaluation after policy implementation [5–7].

In the classification of sports industry policies, the classifications given by researchers in different research periods according to the development of the sports industry at that time are different. The research on sports industry policy is still relatively comprehensive. Whether it is from the overall sports industry or the subindustries under the classification, after more than 20 years of research, significant results have been achieved [8]. Different regions pay different attention to sports industry policies. The factors considered in formulating policies are closely related to the environment in which they are located. Geographical location, resource advantages, and the social economy will all affect the layout of sports industry policies to a certain extent [9]. Zhang et al. took the sports industry in Shaanxi Province as the research object and analyzed the total scale of the sports industry, the scale of the sports industry, the scale of the construction industry, the scale of the sports industry, and the scale of sports infrastructure practitioners [10].

Since 1994, relevant research on the sports industry policy has been developed, and a lot of achievements have been achieved. The relevant research on the sports industry policy has an important role in the development of the sports industry. Although the current research results in China are relatively significant, the level of research is also deepening. However, the problems existing in the research cannot be ignored [11]. Researchers usually tend to analyze the policy quickly after the country has introduced a certain policy but ignore the actual problems exposed by the policy in the current development process of the sports industry. Among the many studies, there is a lack of research on technical policies, and the multidisciplinary and cross-departmental cooperative research on sports industry policy has not received much attention [12].

In recent years, with the continuous progress of science and technology, China has entered the era of big data. By using various technologies of computer software and combining disciplinary knowledge with computers, significant research results can often be obtained [13]. The knowledge graph is an interdisciplinary scientific research method that relies on the powerful computing functions of computers, combined with the characteristics of applied mathematics, graphics, and other disciplines, the results of which are often the overall situation and structure type of the development of the subject area, which is beyond the reach of traditional research methods [14]. The technical application of a knowledge graph requires data as the basis. Through data import, the imported data is finally presented in the form of a cooccurrence graph. According to the cooccurrence graph, we can clearly know the research hotspots of a certain discipline at different time stages. According to the results, analyze the overall development trend of the discipline and predict the future development prospects. As one of the important research methods in science, knowledge graph visualization analysis is closely related to the scientific field. In 1938, Bemar made an early subject graph. With the rapid development of computer technology, by the middle of the 20th century, Price presented the growth law of scientific knowledge index through curve visualization [15]. Garfield created the "Science Citation Index," which led to research on "document coupling," "coword analysis," "science citation network," and other related issues, and promoted the further exploration of scientific knowledge visualization by scientific researchers [16].

In the field of sports science in China, a knowledge graph is mainly used as a research method, and the functions of cooccurrence analysis, coword analysis, citation analysis, and other functions in a knowledge graph are used to conduct scientific statistics on literature data, so as to find out the research hotspots of the discipline [17]. Its main research model is to start research around the research hotspots, research frontiers, cooperation networks, and other issues of a certain discipline, and finally, to sort out the basic model of the scientific development process and trend of the discipline. Its main coverage includes physical education, sports science, school sports, campus football, aerobics, martial arts, table tennis, physical education teachers, physical education courses, physical fitness training, national fitness, tai chi, basketball, youth, etc. [18].

From the birth of the knowledge map in 2003 to the introduction of the knowledge map into the field of scientific research in China in 2005, it was only two years apart. The rapid development of science and technology and the advent of the information age have provided the possibility for the rapid and widespread dissemination of knowledge. In the early days of the introduction of knowledge maps into China, it was mainly used in the research of library information and digital libraries. With the change of time, the application of knowledge maps in various research fields in China is also increasing, such as education theory and management, computer software and research on its application, higher education, scientific research management, enterprise economics, and other research fields. The earliest use of knowledge graphs in sports research in China was in 2010 when Wang Qi from the Shanghai Institute of Physical Education applied this method to sports related research. Since then, knowledge graphs have attracted more attention in sports research, and this method has been applied to many research fields such as sports consumption, sports training, martial arts, sports policy, and other fields, and the research results have been remarkable. Although in the related research of sports disciplines, there is no systematic research in the field of sports industry policy research.

However, there is almost no research on the overall grasp of the entire sports industry policy research field. Therefore, this study aims to draw a statistical and relevant map of the literature on related research on sports industry policy and provide an overall understanding of China's sports industry policy. The research field is systematically analyzed, and the future research trend of this research field is grasped by careful analysis of the map.

In the following steps, we first introduced research methods which are needed for our study in Section 2. Then, we analyzed sports industry law including cooperation degree and cooperation rate analysis, author collaboration network analysis, cooperative network analysis of research institutions, research hotspots and theme analysis of sports industry policy, cluster analysis of research hotspots, and analysis of research topic in Section 3; finally, we present our main conclusions in Section 3.

2. Research Methods

This paper takes "the current situation of policy research of sports industry law" as the research object and retrieved from the full-text database of China National Knowledge Infrastructure (CNKI) and Web of Science (WoS) academic journals during the twenty-five years from January 1, 1990, to December 31, 2022, a total of 2295 articles Recreational sports related research literature is used as the data of this study, and the data is visualized and analyzed by Citespace6.1.R2 (64-bit) software.

2.1. Literature Method. The literature method is a theoretical foundation for the research content before conducting academic research. By consulting a large number of literature materials and reading, sorting, and analyzing them, researchers can be prompted to have a relatively full understanding of their research directions. There is sufficient theoretical support in the process, which is a research method with important guiding significance for this research. According to the research purposes and tasks, this research has specific plans and measures in terms of literature data collection, as follows.

The first step is to use network database resources such as China National Knowledge Infrastructure (CNKI) and WoS academic journal full-text database, China National Knowledge Infrastructure (CNKI) master and doctoral dissertation database, VIP Chinese journal full-text database, Chaoxing Library full-text electronic books, Wanfang Data Resource System and other network database resources, Extensive collection of relevant research materials on "sports industry policy," "knowledge map," and "visualization analysis"; the second step is to consult relevant books such as " sports industry," "knowledge map," and "scientometrics" through libraries with large collections such as Beijing Library, Peking University Library, and Tsinghua University Library; the third step is to inquire about the policies and regulations related to the sports industry through the Internet platform. In order to ensure the authenticity of the inquired results, a review is carried out on the release platform, mainly based on the official website of the government agency.

The above three steps can effectively help us to obtain a large number of sports industry research materials and then organize, classify, and analyze these materials so as to lay a solid foundation for the next research.

2.2. Knowledge Graph Method. The knowledge map method is a relatively advanced data analysis method in the field of scientometrics. Its main development advantage lies in the combination of traditional metrology and modern information technology and cross-analysis by integrating graphics, applied mathematics, and other disciplines to present a more intuitive and clear visual map. Through the review of relevant literature, it is found that the relevant research on knowledge graphs in China started in 2005. In the field of sports science, there is relatively little literature using the knowledge graph method for research, and there are no articles that use the knowledge graph method to visualize the current status of my leisure sports research [19].

According to the related research on the knowledge graph method, the commonly used knowledge graph analysis software programs in the current academic research field are Bibexcel, Citespace, SPSS, Ucinet, TAD, etc. Sources also have different degrees of applicability. Therefore, according to their relevant characteristics, this paper chooses the Citespace software system that has a higher degree of data fit and more powerful analysis data of China National Knowledge Infrastructure (CNKI) and WoS for research.

2.3. Word Frequency Analysis Method. Term frequency analysis method refers to TF-IDF (term frequency-inverse document frequency) term frequency-inverse document frequency, a common weighting technique mainly used for data retrieval and text mining, used to measure the effect of a certain word on an article or a document. The importance of the field literature data, the importance of the word to the literature in which it is recognized by the frequency of occurrence of the word, and the relationship between the two is proportional. The higher the frequency, the greater its importance and the most relevant research [20]. For academic research, word frequency analysis can timely reflect the research frontier issues of a certain discipline and understand its development trend. For this study, the word frequency analysis method can reveal the focus and research hotspots of China's sports industry policy research in five development stages Mathematical statistics method.

Mathematical statistics is a branch in the field of mathematics. It is a common scientific research method that uses statistical methods to derive and analyze relevant data to obtain research results. In this paper, the mathematical statistics method is mainly used for the statistics of the authors and the institutions to which the authors belong so as to facilitate the analysis of their relevant cooperation degrees.

3. Results and Discussion

3.1. Cooperation Degree and Cooperation Rate Analysis. The arrival of the information age of human society is accompanied by the rapid development of science and technology, and the level of scientific research is also constantly improving with the growth of science and technology. The objects of scientific research are developing widely, and the scope of research is developing towards high depth and precision. The explosive growth of scientific knowledge and the highly specialized development of scientific knowledge and technology have become an undeniable fact [21]. At present, relatively independent scientific researchers can generally provide limited funding and resources for the scientific research process. Therefore, collaborative research will gradually become the mainstream form of disciplinary research. The forms of cooperation are also becoming more and more diverse, including various forms of cooperation between countries, between scientific research institutions, between authors, and between individuals and scientific research institutions. In the research of sports, it is an extremely important form for researchers to carry out scientific research activities in the form of cooperation. Sports industry policy is a research branch with strong comprehensiveness and professionalism. Through scientific research cooperation in this field, the online analysis and research will explore the general situation of cooperative research in the field of sports industry policy research, which will help scholars to carry out more research in the form of cooperation in the future and promote the in-depth development of sports industry policy research [22].

There are many ways for scientific researchers to cooperate. Participating in research and publishing research results can be understood as collaborative research between authors. The cooperation between different institutions is mainly manifested in the coauthoring of papers or books between different institutions. Understanding the distribution of researchers and scientific research institutions in the field of sports industry policy research and the cooperation in the process of scientific research, and exploring the trend of cooperation among future scholars in this research field are not only conducive to objective scientific guidance in future sports scientific research but also helpful. It is helpful for scholars to choose good partners in the research of sports industry policy and to promote the comprehensive improvement and development of sports industry policy while promoting the research and development of sports industry policy.

The two indicators of cooperation degree and cooperation rate are mainly used to indicate the cooperation scale in the subject category. The larger the obtained value, the higher the cooperation scale or degree of cooperation. Cooperation degree and cooperation rate are two benchmarks for weighing cooperation in scientific research [23]. Cooperation degree is the ratio between the total number of authors of relevant literature and the total number of papers in a certain period of time, and the cooperation rate is the sum of the number of cooperative papers and the total number of papers in a certain period of time. The ratio between the two is affected by many factors, such as the difficulty of research projects, the relevant capabilities of scientific researchers, and the level of scientific research management.

In order to more intuitively understand and grasp the cooperation degree and cooperation rate of the authors of the sports industry policy research, the collated sample data are counted and analyzed, and the authors with the same name in the data are excluded. It is concluded that 3400 authors have published 2295 journals. According to the above formula on the degree of cooperation, it is concluded that the degree of cooperation in this research field in China is 1.5, and the cooperation rate is 45.86%. The relevant statistical results are shown in Tables 1 and 2.

According to Table 2, there are a total of 2,295 pieces of literature on sports industry policy research selected in this study. By importing the data into Excel for relevant statistics, it is concluded that the total number of authors is 3,400, and the degree of cooperation of the papers is calculated to be 1.5. The degree of cooperation in the field shows that the higher the degree of author cooperation, the higher the academic influence of the research, and the two are positively correlated. Compared with the cooperation degree of sports industry research and national fitness research, the cooperation degree of China's sports industry policy research has a large room for improvement. Compared with the degree of cooperation in other research fields, there is a certain gap. First, it is because of the differences in the degree of development between the research fields. Another reason may be the uneven distribution of resources in the field of sports industry policy research. In view of the current low level of cooperation in China's sports industry policy research, research in this field still needs to strengthen cooperation between scientific researchers and improve the level of scientific research management in this field, especially the field of sports industry policy and other research cannot be ignored. There is overlap and integration between fields, and the scientific research methods in sports research are constantly enriched to improve the degree of cooperation between authors.

From the data in Table 3, it can be seen that cooperative research in the field of sports industry policy research has gradually become a general trend. The number of papers written by one person alone accounts for 60.2%, more than half of the total, indicating the scale of cooperation in this field in China. There is still room for improvement. In the coauthored literature data, the number of coauthors with 2 people and coauthors with 3 people occupies an absolute advantage, indicating that the current cooperation scale in the field of sports industry policy research has a large room for improvement. The coexistence of professionalism and comprehensiveness is a major feature of sports industry policy research.

3.2. Author Collaboration Network Analysis. In order to have a more comprehensive and intuitive understanding of the cooperation between authors in the field of sports industry policy research, this study uses Citespace software to draw the author cooperation map of the journal literature on sports industry policy research from Web of Science. The drawing process is as follows: In Citespace software, the time node is set to 2003–2022, the time slice is set to 1, the node type is selected as the author, and the threshold is selected as 50, which means the top 50 authors every 1 years. Selecting Pathfinder (path-finding network algorithm) to cut the map reasonably, and finally obtain a map suitable for this study, and conduct a comprehensive analysis of the corresponding map. Figure 1 shows the cooccurrence map of authors in the field of sports industry policy research in China.

In Figure 1, the connection between different scholars is represented by the connection between different points, and the degree of cooperation between scholars is positively correlated with the width of the connection. The size of the font of the name represents the number of articles published in the field. It can be seen from the map that in the current research on sports industry policy, researchers such as Shcuster, Galea, Henderson et al. are relatively productive

Table 1: Statistics of the top ten articles by citation frequency.

Authors	Research institute	Papers	Source	Citations	Total upload
Feng et al.	Capital Institute of Physical Education	The Equalization of Sports Public Service and Its Fiscal Policy Choice	Journal of Shanghai Institute of Physical Education	170	2680
Bao et al.	General Administration of Sports of the People's Republic of China	Theoretical Thinking on Establishing and Perfecting the New National System	Journal of Tianjin Institute of Physical Education	190	1344
Chen et al.	Wuhan Institute of Physical Education	Research on the Policy Choice of China's Sports Industry Structure and Industrial Layout	Sports science	153	4381
Zhao et al.	National Sports Commission Research Institute	Research on the Policy System of Sports Industry	Sports science	182	1693
Don et al.	Department of Education, Science and Culture, Ministry of Finance	Status quo, development measures and experience of foreign cultural industries	Economic Research Reference	107	3132
Li et al.	Department of Sports Economics, General Administration of Sports of the People's Republic of China	The Present Situation and Thinking of the Development of China's Sports Industry	Sports market	165	498
Yu et al.	Shanghai Institute of Physical Education	Research on the Development Strategy of China's Sports Industry	China Sports Technology	102	1642
Lei et al.	Beijing Sports University Venue Office	General situation and existing problems of sports venues development (summary)	Shandong Sports Technology	160	1322
Xiong et al.	Beijing Sports University, Jiangxi Normal University	The Formation, Evolution and Reconstruction of China's Competitive Sports Development Model	Sports science	212	4090
Li et al.	School of Physical Education, East China University of Technology	The Enlightenment of "Supply Side Reform" to the Development of China's Sports Industry——Based on the Perspective of New Supply Economics	Journal of Wuhan Institute of Physical Education	4	5633

Table 2: Statistical table of cooperation degree of sports industry policy research literature.

Total number of papers	2295
Number of authors	3400
Author cooperation	1.5

Table 3: Statistical table of coauthorship of research papers on sports industry policy.

Number of authors of a single article	Number of papers	Contribute %
One person	1250	54.4
Two persons	445	19.4%
Three persons	100	4.3%
Four persons	30	1.3%
Five persons	25	1.1%
Above 6 persons	5	0.2%
Total	2295	100%

authors. The degree of cooperation between the research teams can be seen on the map. The degree of cooperation among the research teams, such as the Gani et al. team, is relatively high. The Schuster team's research in the field of sports industry policy research is involved in the policy sector of the sports service industry, mainly conducting research on the policies of the sports service industry. Its research focuses mainly on the overall environment of the sports service industry, the basic characteristics of policies, the evaluation system of policy changes and implementation effects, etc., and the research results are quite rich.

Through the analysis of the map combined with the relevant status quo, it can be seen that the degree of cooperation between the authors of sports industry policy research needs to be improved. The team research is mainly inclined to carry out in-depth and detailed research on a specific sector, and only a small number of research teams are involved in relatively extensive research content. Because the number of high-level research teams is limited, there is still room for improvement in the excellent results of cooperative research on sports industry policy research.

3.3. Cooperative Network Analysis of Research Institutions. From a comprehensive perspective, analyze the scientific research cooperation and scientific research achievements of research institutions in the field of sports industry policy research and analyze the role of each research institution in the cooperation network in this research field, which is conducive to scientifically and objectively exploring sports industry policies. This study uses Citespace software to draw the cooccurrence knowledge map of research institutions, which can clearly show its specific distribution and try to use the cooperation network map of scientific research institutions in sports industry policy research literature and combine the corresponding literature data. The method of drawing the cooperation map of scientific

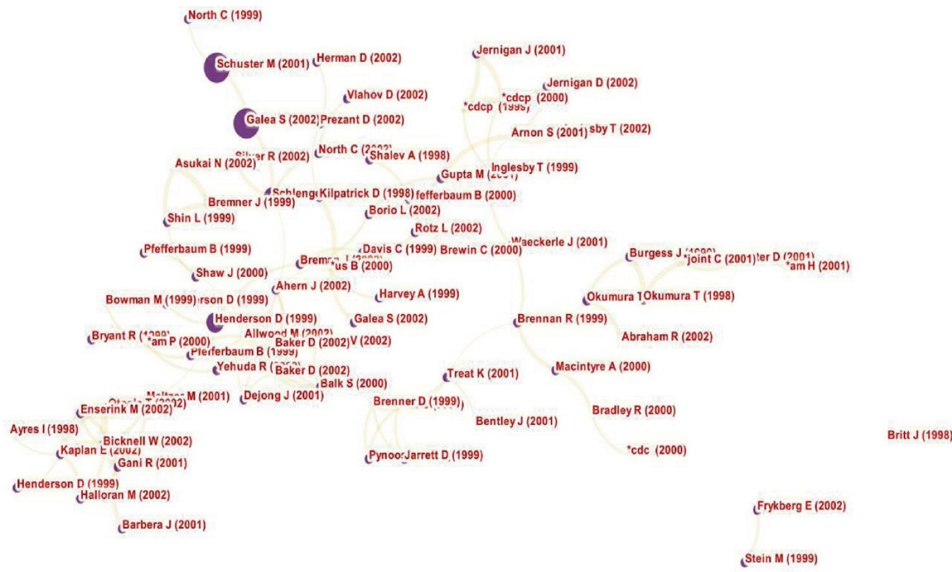


Figure 1: Author cooccurrence map for sports industry policy research.

research institutions in the field of sports industry policy research is the same as the drawing method of the author's cooperation network map, and the institutional cooperation network map of sports industry policy research is obtained.

From Figure 2, it can be clearly understood that the main force of sports industry policy research comes from colleges and universities, mainly concentrated in professional sports colleges or sports colleges among comprehensive colleges, although other types of schools are also involved in the sports industry. But policy research's status is not obvious. There are 10 connections between institutions, indicating that cooperation between institutions exists, but the intensity of cooperation still has a strong room for improvement in the future. It can be seen from the map that the main areas of cooperation between institutions are different colleges and universities in the same locations or colleges and universities in different cities in the same region. Harvard University, Columbia University, and other scientific research institutions are the main research institutions in this field, and their publication volume also ranks at the forefront of all research institutions. These institutions can be considered as knowledge creation and distribution centers for sports industry policy research, and these institutions have strong research strength. Among these institutions, Havard University, in particular, has conducted in-depth research on sports industry policy, especially the number of published papers ranks among the top of all research institutions, and its research scope is also relatively extensive, mainly in sports industry policy, sports industry development-related policies, which also has rich research results in the fields of sports rule of law and policies related to sports venues. From the perspective of the connection of institutional cooperation, there are strong and weak alliances in the current cooperative institutions of sports industry policy research, such as the cooperation between Harvard

Figure 2: Institutional cooperation network map of sports industry policy research.

University and Columbia University's School of Physical Education, which is conducive to promoting the improvement of teams with weak research strengths. There are strong alliances, such as the cooperative research between North University Penn and Uniformed Ser University, which can produce high-quality and efficient research results. However, it can also be seen from the map that the weak link in the current cooperation between sports industry policy research institutions is that the scale of cooperation is small, and the density of cooperation and the external cooperation between regions is also relatively small. The cooperation between the eastern, central, and western regions is mainly concentrated in the eastern region. The cooperation between the central and eastern regions is better than the cooperation between the western and central regions, while the cooperative research between the western and central scientific research institutions is hardly carried out.

3.4. Research Hotspots and Theme Analysis of Sports Industry Policy. The important programmatic vocabulary that reflects the subject content and research focus of the article is the keyword. Drawing a keyword cooccurrence map and analyzing it makes it possible to intuitively understand the hot content and topic distribution in this research field. Counting the keywords in the field of sports industry policy research and drawing a knowledge map can reflect the changes between the research hotspots in different periods. As the refined vocabulary of an article, keywords can intuitively reveal the research theme and focus of the paper. To a certain extent, keywords have obvious timeliness and freshness and can reflect the author's cognition of the research field.

Citespace software comes with the function of counting the frequency of keywords. It can clearly show the frequency and clustering relationship in a visual way on the basis of counting the occurrence frequency of related words. In order to show the hotspots and research topics of sports industry policy research clearly and intuitively, this study used Citespace software to draw the keyword cooccurrence map of the keywords in the sample data for the 2295 articles collected in this field. The method is the same as that of drawing the author's cooperation network, but there is a difference in selecting the node type. After the graph is properly trimmed, the obtained graph is shown in Figure 3.

In the keyword cooccurrence map of sports industry policy research, there are a total of 300 nodes and 410 connections. The differentiation of the analysis objects is displayed by different nodes, and the frequency of keywords is determined by the size of the nodes. The connection between keywords is represented by the thickness of the connection between the nodes, and the two are positively correlated.

From Figure 3, it can be seen that the keywords with larger nodes are sports industry, sports industry policy, sports economy, policy, sports policy, industrial policy, development, sports service industry, supply-side reform, and other keywords. It shows that the hotspots of sports industry policy research are mainly concentrated in the above-related sectors. It can be seen from the cooccurrence map that the lines between different keywords are generally thick and thin. In the map, the keyword sports industry has the largest node, mainly because the research on sports industry policy is originally carried out around the relevant policies released by the sports industry. Secondly, the subject search is used when retrieving documents, and the subject words are limited between the words "sports industry" and "policy." Therefore, among the relevant documents searched, the word "sports industry" has the largest node in the graph.

3.5. Cluster Analysis of Research Hotspots. Using the clustering function of Citespace software, through the analysis of keyword clusters, select the first 10 clusters and interpret the theme trends of sports industry policy research through these eight clusters, as shown in Figure 4.

Citespace software has the function of a synthetic cluster analysis model, which can divide keywords into multiple clusters according to the cooccurrence relationship between keywords, and each independent cluster or closely related adjacent clusters represents similar research topics. This study uses the clustering function of Citespace software, selects the top 10 clusters through the analysis of keyword clusters, and uses these eight clusters to interpret the theme trends of sports industry policy research, as shown in Figure 4.

The first cluster group is sports industry. The sports industry has not been around for a long time, but its development has been accelerating over time. At present, the development of the sports industry has received unprecedented attention, and related research on the sports industry has developed rapidly in recent years.

The second clustering group is countermeasures. Countermeasure research is prevalent in the research of various disciplines. In the field of sports industry policy research, researchers often conduct research on relevant policies on the basis of the current development of the sports industry so as to provide substantive and effective measures for the development of various industries in the sports industry. Operable countermeasures and countermeasure research can often be more realistic in testing the practical application of scientific research results.

The third cluster group is sports industry development. From the beginning of the development of the sports industry to the present, the research field of the sports industry has never stopped according to the development of the sports industry in various periods, so as to study its current situation, which is of great reference significance for grasping its future development trend. In the research of sports industry policy, it is particularly important to grasp the current situation of the development of the sports industry so as to match the actual development of the sports industry with the relevant policies issued and explore the key direction of the future development of the sports industry.

The less the cluster number, the more articles in that cluster. Therefore, we only analyzed the top three clusters, which can be the best illustrate research hotspots.

3.6. Analysis of Research Topic. "Research hotspots" are the focus of scientific researchers on certain special issues within a certain period of time. One of the samples that can best reflect the research topic in the literature is the keyword, and when this research retrieves relevant journal literature on sports industry policy research, the search criteria is selected as the topic, and the subject heading is limited to "sports industry" and contains "policy." Therefore, the keywords of the literature collected in this way can better represent the research hotspots in this field when conducting data sample statistics and analysis. According to the drawing of the knowledge map of

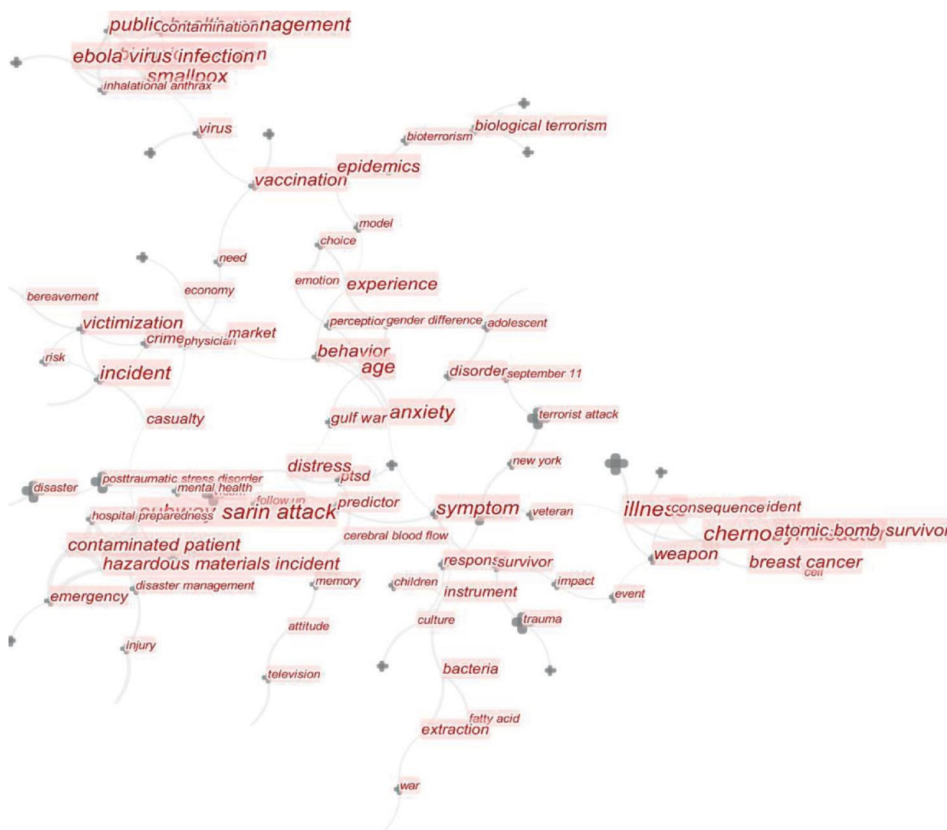


Figure 3: Keyword cooccurrence map of sports industry policy research.

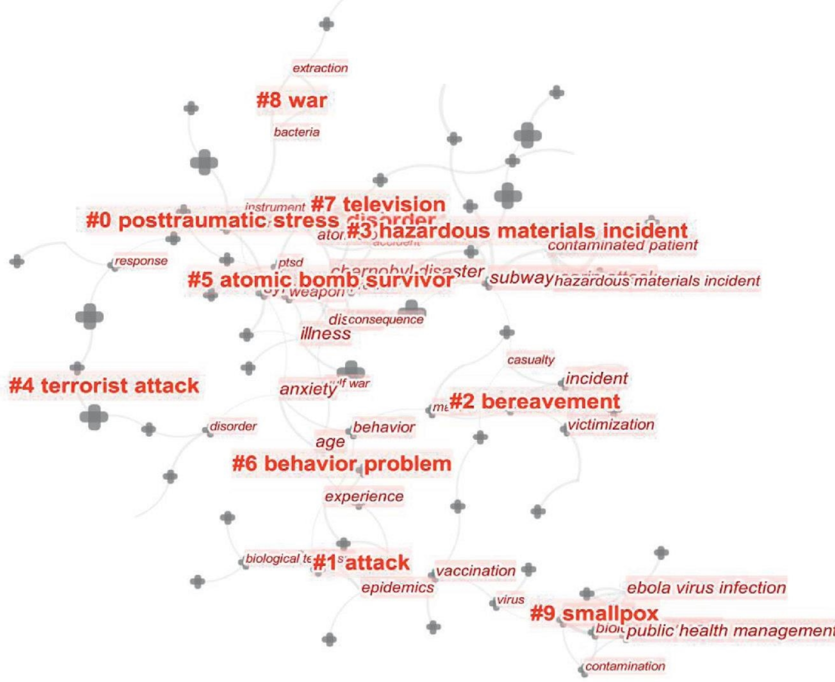


Figure 4: Sports industry policy research keyword clustering map.

keyword cooccurrence in the research of sports industry policy by software, this study makes statistics on the top 20 high-frequency keywords, as shown in Table 4.

From the relevant data in Table 4, it can be seen that "sports industry" is the most frequently occurring keyword in sports industry policy research because the research on

Table 4: Statistics table of top 15 high-frequency keywords in sports industry policy research.

Ranks	Keywords	Frequency	Centrality	Emergence rate
1	Sports industry	320	1.1	
2	Policy	70	0.2	10.1
3	Sports economy	51	0.2	
4	Development	50	0.1	5.2
5	Sports industry policy	40	0.1	
6	Sports	40	0.08	7.7
7	China	30	0.14	
8	Measures	30	0.07	
9	Industry measure	27	0.08	5.5
10	Sports policy	20	0.01	3.3
11	Industry	20	0.04	
12	Sports industry development	17	0.02	
13	National fitness	15	0.0.9	4.00
14	Question	15	0.01	
15	Sports tourism	13	0.02	4.0

sports industry policy will inevitably lead to research on the sports industry. From the literature review of knowledge graph-related research, it can be seen that betweenness centrality can measure the importance of nodes in the network. A measure of the importance of a node in Citespace is betweenness centrality, which refers to the ratio of the shortest path passing through a point and connecting the two points to the total number of shortest path lines between the two points in the network, which is used to express the degree to which a node controls the interactions between others, and the position and importance of an individual or organization in its network is represented by it. From Table 4, the intermediary centrality of each keyword shows that the sports industry occupies a considerable proportion of the sports industry policy research and which is an indispensable and important part of the sports industry policy research field.

The importance of the sports industry continues to rise. In recent years, the state has frequently issued policies on the development of the sports industry and even issued plans for the development of a single sport. As a result, the proportion of the sports industry in the national economy has continued to rise, and the composition of the national economy has been optimized. Policy, sports economy, development, sports policy, etc. are the keywords with the highest frequency in sports industry policy research, indicating that the theme of sports industry policy research mainly revolves around the sports industry, the development of the sports industry, and its related supporting functions and has a supporting role the policy of conducting research. The high frequency of occurrence of keywords indicates that these have become the focus of scientific research workers in the research of sports industry policy because the research of sports industry policy is inseparable from the sports industry and even more inseparable from related policies. At the same time, the continuous development of the sports industry is the research possible to continue. From the perspective of the sports industry, the current attention the sports industry has received is inseparable from the intensive introduction of relevant sports industry policies by the state.

4. Conclusions

This paper selects 2,295 journal documents on sports industry policy research from 1990 to 2022 as the sample data for this research and used Citespace software to visualize and draw the relevant content for analysis. The main conclusions are as follows:

The annual output of journal literature on sports industry policy research and its corresponding time has its own uniqueness. In 2008, the number of related research literature began to show a rapid growth trend, and then the output of papers showed a stable development trend. Colleges and universities are the main institutions for sports industry policy research. Among them, sports colleges in comprehensive universities and professional sports colleges occupy a dominant position in research results. The research strength is mainly concentrated in the eastern and central regions, and the core authors are mainly concentrated in the central and eastern regions.

The cooperation situation of educational, industrial policy research is gradually taking shape, but the scale and degree of cooperation in this field have a large room for improvement. Concentrated in the central and eastern and a few cities in the west, the distribution of research in this field by scientific research institutions is uneven, and the regions are mainly concentrated in the central and eastern regions, which also shows the current regional development differences of the sports industry.

Through the statistics and mapping of relevant data, it is found that the research hotspots of sports industry policy mainly include sports industry, sports economy, development, countermeasures, sports industry policy, sports consumption, national fitness, and other research hotspots. The content of sports industry policy research is based on the current research status. The two research hotspots of the sports industry and sports industry policy have always been accompanied by research in this field. Because of the particularity of the topic selection, the research hotspots are definitely inseparable from these two points. Compared with the persistence of the first two research hotspots, sports economy and countermeasures appeared a little later. The research hotspots of sports consumption and national fitness have only rapidly become the research hotspots in this field in recent years.

The hotspots of sports industry policy research are differentiated according to different time periods. The future trends of research mainly include interdisciplinary research and cooperative research between different regions will gradually become a trend, cooperative research between authors will become a common phenomenon, and the distribution of research subjects will be more uniform, and more comprehensive colleges and other research institutes will join.

This research will provide a clear research context for future sports industry policy research and provide new perspectives and new ideas for grasping the hotspots and future trends in the research process in this field.

In the future, we should break the geographical limitations of research, strengthen the cooperation between research teams, and encourage strong alliances between research forces, but we should not ignore the exchange and collaboration between talents. The society can create a good communication platform for each research team, promote the balanced development of the main body of sports industry policy research, and thus promote the continuous development of China's sports industry. In particular, the cooperation between cross-regional research teams and the cooperation between comprehensive research institutions and professional research institutions will play a good role in promoting the comprehensiveness of sports industry policy research. Cultivate multidisciplinary talents, especially the research teams of comprehensive colleges and financial colleges, and at the same time, professional research institutions in the field of the sports industry cannot be ignored.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. Maedche and S. Staab, "Ontology learning for the semantic web," *Intelligent Systems IEEE*, vol. 16, no. 2, pp. 72–79, 2001.
[2] Y. Yang and Z. Liang, "Research on the present situation of public sports products supply in rural areas," *Proceedings of the 2nd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE*, vol. 205, pp. 1102–1105, 2018.
[3] Yu Zhang, J. Ding, Y. Hui, M. He, W. Wang, and F. Xiong, "A study of the influence of collaboration networks and knowledge networks on the citations of papers in sports industry in China," *Complexity*, vol. 2022, Article ID 9236743, 10 pages, 2022.
[4] L. Tang, "Study on coupling coordination degree between regional sports industry development and healthy China construction," *Journal of Healthcare Engineering*, vol. 2022, Article ID 7785267, 5 pages, 2022.
[5] L. Yuxin and L.2 Lu, "Countermeasures for the sports industry to promote the construction of China's economic internal circulation," *International Journal of Educational Management*, vol. 6, no. 4, 2021.
[6] D. Garewal Das, "Research on sports industry governance mode," *Journal of Social Science and Humanities*, vol. 3, no. 8, 2021.
[7] X. Ye and Li Chen, "The current trend of structural change in sports industry in China is studied," *International Journal of Education and Economics*, vol. 4, no. 1, 2021.
[8] J. Torres Pruñonosa, M A. Plaza Navas, D. Martín Francisco, and P. R. Camilo, "The sources of knowledge of the economic and social value in sport industry research: a Co-citation analysis," *Frontiers in Psychology*, vol. 11, 2020.
[9] Y. Song and Ye Cheng, "Exploration on the conversion path of the new and old kinetic energy of Chinese sports industry," *Frontiers in Sport Research*, vol. 2, no. 60, 2020.
[10] J. Zhang, Li Nan, and X. Yang, "The development status and countermeasures of the sports industry in Shaanxi Province," *Frontiers in Sport Research*, vol. 20, no. 50, 2021.
[11] Li Ka-sing, "Research on the development of sports industry structure in China[J]," *The Frontiers of Society, Science and Technology*, vol. 2, no. 9, 2020.
[12] Research and Markets and Global Sports Industry, "COVID-19 to cause serious financial and logistical implications - ResearchAndMarkets.com," *Medical Letter on the CDC & FDA*, 2020.
[13] C. Deng, Z. Tang, and Z. Tang, "Historical opportunity and structural transformation of sports industry development under the background of new information technology," *Journal of Physics: Conference Series*, vol. 1575, no. 1, Article ID 012170, 2020.
[14] Y. Li, "Study on the evolution of governmental relationship of Chinese sports industry policy from the perspective of policy documentation quantification," *Frontiers in Social Science*, vol. 10, no. 07, 2021.
[15] M. S. Kim, S. W. Oh, J. W. Lim, and J. W. Han, "Social big data analysis of demands for sport industry policies," *Korean Journal of Sport Management*, vol. 22, no. 2, pp. 45–66, 2017.
[16] S.-Il Hong, S.-Ki Ahn, and T.-W. Chung, "Examination of determinant factors influencing local governments' sport industry policy adoption and diffusion: a case of major league baseball host cities," *Korean journal of sport management*, vol. 18, no. 6, 2013.
[17] S. K. Kim, "A study of sports industry policy in China," *The Korean Journal of Physical Education*, vol. 51, no. 5, 2012.
[18] X. Zhan, "From budding to developing: The history of china's sport industry policy," *The International Journal of the History of Sport*, vol. 33, no. 18, 2017.
[19] Sustainability Research, "Recent findings from K.H. Hu and Co-researchers yields new information on sustainability research (evaluating the improvement of sustainability of sports industry policy based on MADM)," *Ecology Environment and Conservation*, 2016.
[20] C. Sun, "The evolution and enlightenment of school sports policy implementation in developed countries," *Scientific*

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9841575, 1 page
https://doi.org/10.1155/2023/9841575

Retraction

Retracted: Art Product Recognition Model Design and Construction of VR Model

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Wu, "Art Product Recognition Model Design and Construction of VR Model," *Security and Communication Networks*, vol. 2022, Article ID 3994102, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3994102, 13 pages
https://doi.org/10.1155/2022/3994102

Research Article

Art Product Recognition Model Design and Construction of VR Model

Tianyou Wu

Industrial Design Department, School of Mechanical and Power Engineering, Guangdong Ocean University, Zhanjiang 524000, China

Correspondence should be addressed to Tianyou Wu; yoyo576@gdou.edu.cn

Received 12 April 2022; Revised 23 May 2022; Accepted 1 June 2022; Published 22 June 2022

Academic Editor: Chin-Ling Chen

The artwork embodies a profound human history and carries the essence of human civilization. Its content is complex and covers a wide range. How to use advanced technology to quickly and accurately classify and retrieve is an important research topic in the field. In our study, we first according to the requirements of practical application scenarios and existing data conditions proposed an overall scheme of artwork identification and retrieval. Through the functional analysis of the software required and the comparison of various databases, we present the system architecture design and data conceptual design, and complete the system-level planning and design. Then, the crawler grabbing process is designed to obtain artwork graphic data, the artwork dataset production process and labeling status required for the target application scenario were introduced, and the category imbalance state of the target dataset was analyzed. Moreover, the database table structure design of the artwork identification and retrieval system, design and development of each functional module of the server, and the web client was introduced. Finally, according to the organization, structure, and characteristics of virtual reality system, a product design evaluation system based on virtual reality technology was constructed. A theoretical model VR-PDES was designed for the application of virtual reality technology in product design evaluation. The results of this research are of great significance for people to search for images of unknown artworks and improve the service capabilities and service levels of scenic spots.

1. Introduction

The artwork embodies a profound human history and carries the essence of human civilization. Its content is complex and covers a wide range. How to quickly and accurately classify and retrieve art products with the help of advanced technology is an important research topic in the field. Using the convenience of mobile terminal to obtain image data, combined with the recognition and retrieval technology based on image content, this “Internet +” method can quickly form the actual application effect in the scenic area, making the educational value, cultural value, and even collection of artworks. Value can flow directly and truly to the general public.

Google, Microsoft, Baidu, Hikvision, Taobao, Tencent, and other large domestic and foreign companies are at the forefront of the research and application in the field of imagery. Uber, DiDi, SenseTime, Megvii, and other emerging visual technology companies and domestic co-operation with foreign universities to explore the application of images in security, driverless, retail, and other fields. At this stage, the number of computer vision papers published by domestic and foreign institutions and enterprises in top conferences such as CVPR and ICCV accounts for a large proportion, and the number of papers is also increasing rapidly. Image recognition and retrieval are the areas that researchers focus on.

Identifying and retrieving artworks based on image features mainly involve modeling the image content of artworks, and then identifying and retrieving them based on the representation of the constructed image content. Image recognition refers to giving the category information of images at the semantic level. In the field of artwork, image recognition needs to give the specific category of the image in its field or judge that it does not belong to any category in the field. This multi-classification problem usually has many feasible solutions. Common methods include k-nearest neighbors, support vector machines, adaptive boosting methods, neural network methods, etc. Whether it is a recognition task or a retrieval task, the models they used are based on the feature quantities describing the image content to achieve the ultimate goal. Early information retrieval is mainly based on the content of text annotations, and image retrieval based on text annotations is also one of the most common image retrieval methods. Image visual information is closer to the objective information description of objects than text annotation content. Using image content to identify and retrieve artworks has the advantages of accurate, comprehensive, and objective information. The image content description of artworks can be based on local features or deep convolutional features. Both types of features have received more research attention in the past ten years.

There are many kinds of features that describe the visual content of images, and the first to achieve better results in recognition and retrieval tasks is the local features of the image. The local feature uses the gradient statistics of the region around the stable extreme point of the image to describe this sub-region. The most typical is the SIFT feature proposed by Lowe in 1999 [1]. This feature uses the Gaussian difference and downsampling method to establish a Gaussian blur map and Gaussian difference map in a continuous scale space similar to a pyramid structure and then uses 26 neighborhoods. The extreme values screen out the stable regional extreme points and, finally, use the direction-normalized regional gradient statistics to represent the feature points and their neighborhoods. When using local features for image recognition or retrieval, researchers mainly draw on the research results of document classification and retrieval. Sivic first introduced the bag-of-words model into the image field in 2003 to quantify the local features of images into visual images with certain semantic attributes. Vocabulary forms a visual bag-of-words model BOW [2].

The convolutional features of images are trained from convolutional networks. Important progress has been made in image recognition, object detection, semantic segmentation, and image retrieval. It is a hot research hotspot and application direction. From a cognitive point of view, convolutional neural networks simulate biological cognition to learn the features of input images. The features generated by this learning process are closer to human perception than artificially designed local features. In recent years, a breakthrough for deep convolutional network learning has been made. In 2012, Alex showed the huge advantages of convolutional neural networks by using the leading edge of deep convolutional neural networks in the recognition task of ImageNet large-scale datasets [3]. In the tasks related to image content recognition, classification models such as the VGG model [4], the inception model [5], and the ResNet model [6] continuously refresh the recognition accuracy to new heights in the classification tasks of the benchmark datasets. The size of the parameters is also gradually reduced. In the target detection task of image content, the RCNN model [6], the Fast-RCNN model [7], and the Faster-RCNN model [8] have successively achieved staged progress in detection accuracy and real-time performance. In the pixel-level semantic segmentation task, the FCNN model [9] and the mask RCNN model of Kaiming in 2017 [10] have continuously pushed the accuracy and real-time performance of target pixel-level semantic segmentation to a higher level. In image content retrieval tasks, more and more researchers turn their attention to deep convolutional models.

Regardless of whether local features or deep convolutional features are used, when training images for classification tasks, the problem of dataset category imbalance often occurs. This type of problem is essential that real-world data conditions impose constraints on model recognition performance. Therefore, image data category imbalance is a common problem that classification models in pattern recognition need to face. Faced with this kind of problem, researchers have tried a variety of coping methods and techniques in the past ten years, mainly at the dataset level and the model level, to reduce the adverse effects of imbalanced datasets. When evaluating the recognition performance of the model, in order to more objectively evaluate the classification status of the classification model for each category in the imbalanced dataset, the researchers generalized the ROC curve and AUC value commonly used in classification problems in the medical field to common application scenarios to evaluate the performance of the model. For example, in 2006, Fawcett gave a systematic introduction to ROC curves and AUC values [11].

There are two main methods of image recognition at present: one is to construct visual words to establish image representation for identification and classification, and the other is to construct neural network models (mainly convolutional neural network models) for identification and classification. In 2003, Sivic proposed the bag of visual word model (bag of visual word, BOW) and used it to describe the visual content with specific attributes in the image dataset [2]. The class center points formed by local features through clustering can be used as visual words, and a certain dimension of high-level features of deep convolutional networks can also be used as visual words. Looking back at the SIFT feature extraction process summarized by Lowe [1], the process can be divided into two steps: finding key points and building local descriptors. In the first step, the key points are formed by extracting the extreme points in the scale space for screening. When performing local feature clustering, a random k-d tree approximation K-means algorithm is often used [12]. The visual word is the basis of the visual content of the image, and the description of the image content constructed by local features is based on this visual word. Since the common visual words between images are difficult to describe the characteristics of a single image, the statistics of such visual words have little contribution to the description of the image content and should be distinguished from the less common visual words between images. When describing image content frequently, visual words are also given different weights. This weight (inverse document frequency, IDF) is inversely proportional to the frequency of visual words appearing between different images. The weighted description is the word frequency-inverse document frequency value, which can accurately describe the image content. After the image content description amount is constructed, it can be recognized and classified by common classifiers, or it can be directly used for retrieval tasks. Therefore, the image content description amount has important content representation significance for image recognition retrieval [13].

There are many choices of common neuron activation functions, such as sigmoid function, tanh function, ReLU function, etc. [14]. These activation functions can approximate the linear or nonlinear output characteristics of biological neuron activation. Multiple neuron structures are connected to form an artificial neural network. Neurons at various levels in the network can represent signal patterns in different levels of meaning. The learning process of simulating biological neural structure adjusts the weight parameters and bias parameters of neurons, so that different neurons have different responses to the input; that is, neurons can describe various input patterns, and neurons located in the same layer are different from each other. The larger the output value, the more obvious the pattern is. The last layer of the neural network is usually the output layer. The larger the output of the neuron in the output layer, the more obvious the category attribute is, so as to realize the classification. In an artificial neural network, the input of each neuron includes the output of all neurons in the previous layer, which is fully connected. When the neurons in a layer are fully connected, a fully connected layer is formed. However, image signals belong to two-dimensional signals (such as grayscale images) or three-dimensional signals (such as RGB images). The spatial relationship within the image is different from that of ordinary one-dimensional signals. Its texture, color, brightness, and other characteristics often show regional distribution, strip distribution, or linear distribution, and these characteristics can be observed from the operation results by performing a convolution operation with a specific convolution kernel. Therefore, the full connection of multi-dimensional signal input neurons can be realized within the field of view of the convolution kernel, and the basic neurons of the convolutional network can be constructed. The neurons of the convolutional layer only connect the input within the field of view of the convolution kernel, and a convolution kernel only needs one bias parameter. Multiple convolutional layers are stacked to form a deep convolutional network, and the parameter size of the same input is much smaller than that of the fully connected layer [15]. Pooling the convolution output graph by sub-regions (e.g., mean sampling, maximum sampling) can significantly reduce the total number of output neurons and enable the convolution of the next layer to cover a wider initial graph, and the amount of computation is also significantly reduced; usually, a convolutional layer is followed by a pooling layer. The output of the depthwise convolutional layer is passed through the fully connected classification layer; that is, the final category output can be obtained.

The application of virtual reality technology in product design and development provides a new way for product design [16]. Combined with equipment stereo glasses, helmets, data gloves, trackers, etc., and projection equipment through digital models, the virtual world of the product is generated. This virtual world is the combination of the entire virtual environment and a given simulation object. It acts on people through vision and touches to create an immersive feeling. Everyone's operation and modification can timely reflect on the digital model. In this way, the information interaction among people and between people and machines is more real and accurate, and the validation products provide a new development direction for design evaluation [17].

The equipment required by virtual reality technology is complex and expensive. Evaluation data can be collected on the basis of virtual reality technology under current conditions [18]. Use personal computers and software to cooperate to carry out product design evaluations. Process data information through a personal computer and output the final data through the personal computer. Finally, form an intelligent product design evaluation system based on virtual reality technology support, called VR-PDES (virtual reality-product design evaluation system). The application of a VR-PDES for product design evaluation can simplify and intelligentize the product design evaluation method of complex systems. Using the VR-PDES can obtain more accurate evaluation results. It is convenient and practical. It can reduce a large number of mathematical analysis tasks in traditional design evaluation. It greatly shortens the time of data processing. It is convenient and intuitive to obtain product design evaluation results, so as to better assist designers and producers to make correct decisions, improve the efficiency and success rate of product design, and reduce the risk of new products.

In our study, we first proposed an overall scheme of artwork identification and retrieval; second, designed a crawler grabbing process; third, designed a development of each functional module of the server and the web client; and fourth, constructed a product design evaluation system. Section 2 introduced the scheme design method and data collection process of artwork recognition retrieval; Section 3 is the results of the study and discussion; Section 4 is the main conclusions.

2. Scheme Design and Data Collection

2.1. Scheme Design. The identification and retrieval system designed in this paper is aimed at the actual application scene of the scenic spot, using the image content of the artwork as a clue to identify the information needs of the user in the actual scene and the existing graphic data of the artwork in the scenic spot, the graphic data of other collections, and the industry website [19]. The graphic data of the platform and the related graphic data of the e-commerce platform are connected. The starting point of the requirement is the artwork query image submitted by the user, which is also the input query data for the identification retrieval system. The endpoint of the requirement is the most

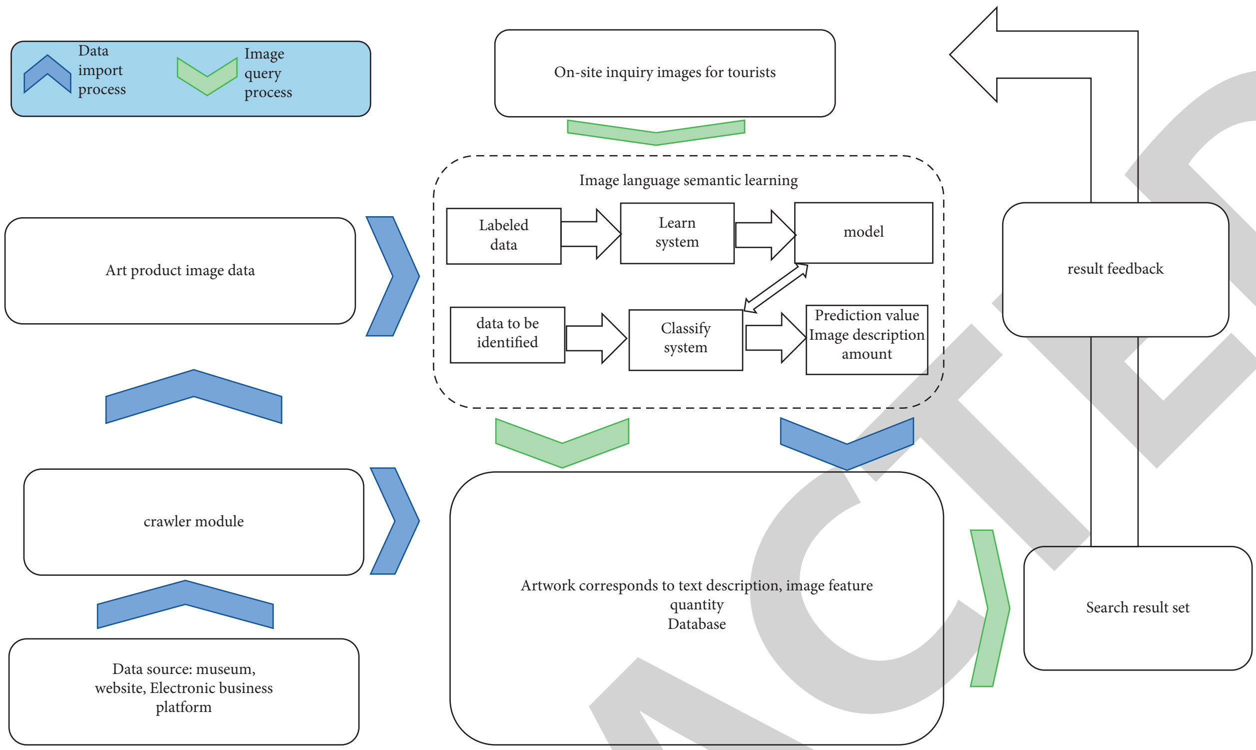


Figure 1: Schematic diagram of the overall scheme of artwork identification and retrieval.

relevant part of the query image in all graphic data that has been imported into the system. The identification and retrieval system must not only understand the images provided by the demand side but also understand all the images included and crawled by the system in advance. Through the understanding of the image, the system will complete the construction of the image description required for the identification and retrieval process, that is, complete the mapping of the image to the image content description (Figure 1). Each image included and crawled by the system has many textual descriptions directly corresponding to it.

The unstructured data of artwork images involved in this paper are acquired in a different way from ordinary sensor data collection processes. Artwork images are mainly captured by a variety of cameras. The sources of the data used in the project include some graphic data of cultural objects provided by partners, as well as graphic data of artworks displayed on third-party websites. Crawling artwork data from third-party websites will effectively supplement the data provided by existing partners. Expanding the data scale through web crawling can not only provide enough information for the query but also provide enough training samples for the identification and retrieval model to make the model more generalizable and effectively reduce the possibility of overfitting.

Due to the diversification of data sources, all the images collected for the first time need to be preprocessed such as de-duplication and outlier point screening, and the existing semantic labels must also be checked by matching to remove erroneous items. Before image semantic learning, the labeled image data are organized to form training datasets, validation datasets, and test datasets. These data will be used for semantic feature learning of artworks and image description encoding learning. The learning process of image semantic understanding can obtain the final recognition model and its model parameters, and the learning process of image description coding can obtain the image coding model suitable for retrieval calculation and storage. After the images of all corpora are encoded by the encoding model, the corresponding encoding description of each image is generated and then imported into the database for use in the subsequent retrieval process. Therefore, the model for image semantic understanding must have sufficient semantic understanding ability and computational response speed. The similarity between the codes created by the image coding model should be consistent with the similarity of the initial image description and have more convenient storage and faster distance calculation feature.

This paper presents the architectural design of the artwork identification and retrieval system, as shown in Figure 2. The entire system is built on the Internet platform, which ensures that every visitor can obtain this artwork identification and retrieval service through image content after accessing the network. The cloud database, SSM framework, and cloud server are the operating platform of this system, providing overall data support and operation support. Above the operating platform is the service layer of the system, including query and retrieval services, data update services, image understanding services, feature extraction service, image coding service, realizing the organic connection between the operating platform, customer service terminal, and management platform. And above the

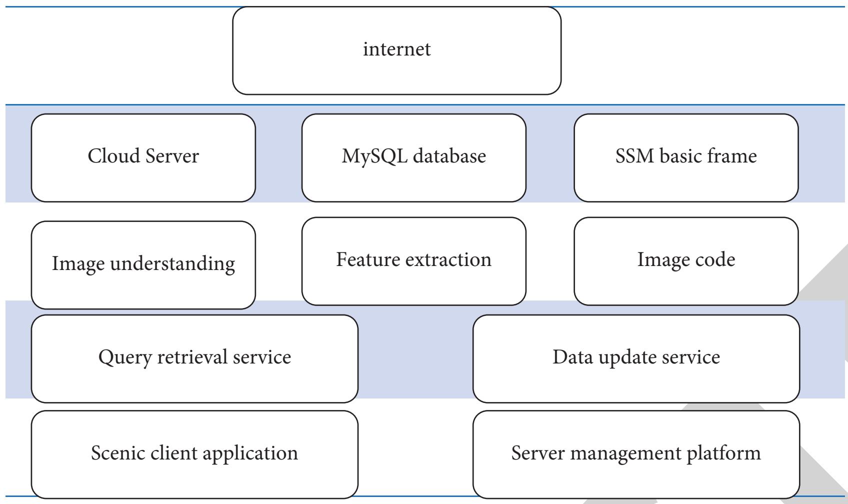


Figure 2: Artwork recognition and retrieval system architecture diagram.

service layer is the application layer directly facing users and administrators, including client applications and server management platforms.

Specifically, the database is an abstract warehouse that stores the target dataset. It organizes and stores the target dataset according to the exact data organization form or the relationship between the data. The development of information technology in the past few decades has formed three common forms of databases: hierarchical databases, network databases, and relational databases [20]. These three databases connect and organize the storage of datasets according to different data structures. There are two main types of database models used in current mainstream Internet applications, namely, relational databases and nonrelational databases. The relational database model reduces various complex data structure relationships to simple binary relationships, that is, data relationships in the form of two-dimensional tables [21]. The relational database implements various operations on the data on the basis of two-dimensional tabular data. One or more relational tables in the data table provide target data link paths for these data operations. Operations such as selection can realize most of the management operations of the database. The birth of nonrelational databases is to deal with the application scenarios of ultra-large-scale massive data and high concurrent requests. Common relational databases are not applicable in such scenarios. However, considering the data scale and the performance of data management, it is more reasonable to use a relational database for the artwork identification and retrieval system in this paper. The earliest relational databases have been around for over forty years.

The relational database developed from theory to today's multiple optional application products is Oracle database, SQL Server, and MySQL. MySQL is different from the other two. It is an open-source database and has good processing efficiency. It is the first choice for small and medium data management systems. MySQL was first used in Linux systems and was gradually ported to other operating systems. It has excellent cross-platform performance. In addition, MySQL occupies less resources and is fast, and the use of MySQL does not require commercial authorization similar to Oracle or SQL Server. Considering comprehensively the performance, operation efficiency, management convenience, and economic conditions, the art recognition retrieval system in this paper adopts the open-source MySQL relational database.

2.2. Data Collection. This research has massive artwork graphic data from various museum websites, industry platform websites, and e-commerce platforms, and e-commerce platforms also have massive artwork images and corresponding market price information data to be supplemented and formatted into a dataset with wide coverage and a large amount of information. This information is suitable for fast crawling by crawlers. The design workflow of crawling is shown in Figure 3.

This research uses Python language to complete the crawler design, also uses the web page parsing tool Urllib library and requests library to crawl static pages, and uses the interface-less browser PhantomJS and driver Selenium to crawl dynamic pages.

Whether using local features or deep convolutional features to build an image recognition model, the supervised training process requires sufficient labeled artwork image data for model learning and training. In the training phase of the model, in order to ensure that the model converges reasonably and determines the performance level achieved by the model, a validation dataset and a test dataset are required for verification and testing.

When crawling artwork data on various website platforms, some artworks themselves have been manually labeled and classified and stored in corresponding page directories according to these categories. These known categories can be used as label data to describe the categories of artworks. After sorting out all category information, removing duplicates and removing errors, there are 30 categories of images corresponding to labeled artworks, and

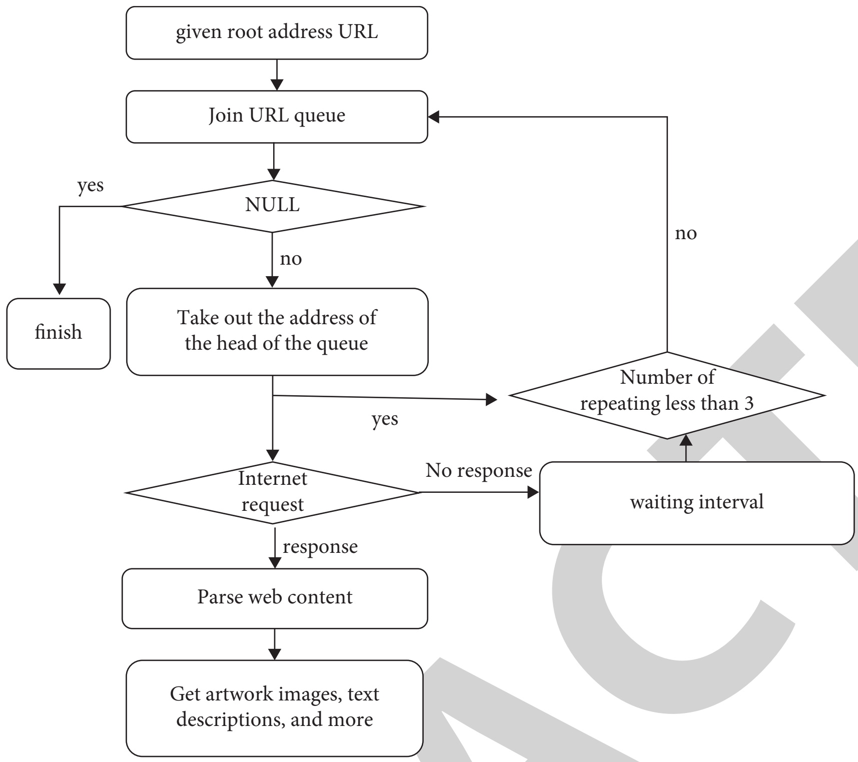


Figure 3: Artwork-related website crawler flowchart.

each category has several subcategories, a total of 29990 labeled image data, and a total of 8,3000 images of artworks without labels. The labeled image data will be used for the training of the recognition retrieval model of the artwork image content. All images with and without tags, as well as the corresponding artwork text and other related information, will be used for the final identification and retrieval application.

The class imbalance problem is particularly pronounced with labeled data, with the highest number of classes having over 3000 data and the lowest having only 40 data (Table 1). Sort the number of all categories from small to large, and draw the class imbalance state diagram as shown in Figure 4. The highest and lowest class number ratio is 83.3, the average adjacent number ratio is 1.2, and the imbalance distribution state is an approximately linear increase of two-stage ladder-like.

Table 1: The initial labeled artwork images divide the dataset by categories.

Categories	Number	Categories	Number
Accessories	60	Jewelry jade	2000
Bamboo and wood teeth	300	Jewelry	80
Books and periodicals	120	Wood carving	230
Bronze	250	Printing/painting	3300
Record player/record	80	Musical opera	100
Ceramic/purple sand	18000	Ancient painting	1200
Clock instrument	300	Disc play	150
Weapon	50	Court religion	200
Weaving	400	Sculpture	400
Enamel	160	Seal	300
Foreign exhibits	60	Old metal coins	100
Furniture	200	Stationery supplies	300
Glassware	50	Bill paper money	190
Law book	500	Gold/silver/ bronze	310
Engraved	300	Life utensils	300

2.3. Construction Method of VR-PDES Model. Kansei engineering provides research method guidance for system construction. “Kansei Engineering” is a comprehensive interdisciplinary subject between art and design, engineering, and other disciplines. Akira Harada, chairman of the Department of Perceptual Cognition and Neuroscience at the Graduate School of the University of Tsukuba and professor at the School of Art and Design, believes that this kind of synthesis and intersection involves many fields of humanities and natural sciences such as art science, psychology, disability studies, basic medicine, and exercise physiology.

Human neural network system provides technical support for the construction of VR-PDES. This system uses artificial neural network method to realize system evaluation data processing. The core algorithm used is network. The idea of realization is to use the design parameters of the product design evaluation in virtual reality as the input end of the network. The design evaluation result of the product is the output end of the network. In the middle is the hidden layer. The product evaluation data and results obtained in the virtual reality environment are used as training samples of the network. The method has a simple operation process,

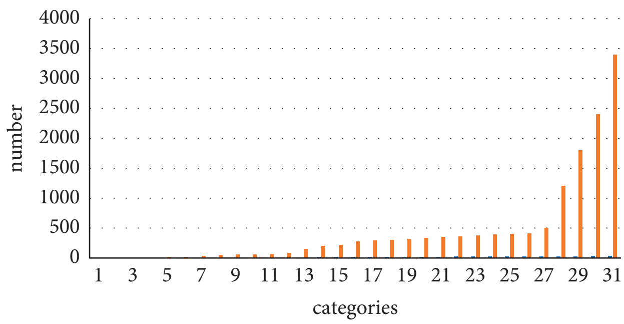


FIGURE 4: Artwork initial labeled data category imbalance distribution map.

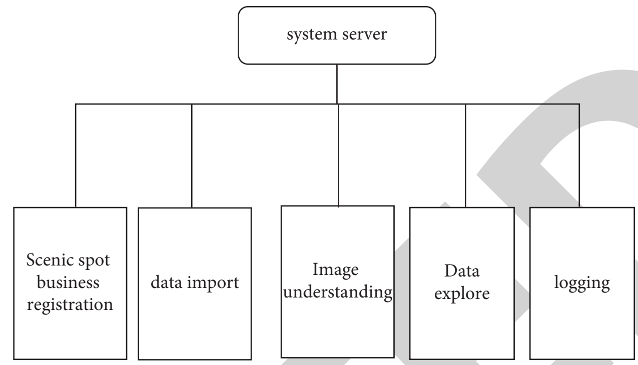


FIGURE 5: System server function tree.

does not require designers to have professional knowledge in multiple fields, and meets the evaluation requirements of general designers.

Because the objective things themselves feel different under different environmental conditions, the corresponding evaluation results and decisions made are also different. Only the design evaluation results obtained in the actual use environment of the product are more accurate and effective, and the decision can be made more correct. There are many uncertain factors in the traditional product design evaluation process relying on the experience and intuition of the evaluator. Therefore, there is a need for a method to reduce the uncertain factors in the comprehensive evaluation process and to evaluate the product design more reasonably. In other words, it is necessary to have a higher problem-solving rate. VR-PDES is a kind of intelligent evaluation system for complex systems that meets the above requirements.

Several key technologies in the construction of VR-PDES are the combination of virtual world, user information tracking collection, and software system. The virtual reality system simulates and generates a virtual reality environment through computer and simulation technology, so that the objects in the virtual reality environment can interact with the user more naturally and realistically. User information tracking and collection are to map the multi-dimensional information of the user's thinking and perception in the virtual reality environment to the digital space of the computer to generate corresponding data information and provide necessary and effective information data for the establishment of the system database.

User information tracking mainly includes key technologies such as spatial tracking, sound localization, visual tracking, and viewpoint sensing, which can help obtain detection and operation data of user operations in virtual reality environments.

High-speed large data processing capability is required between mapping and feedback, so high-performance computing processing technology with high computing speed, strong processing capability, large storage capacity, and strong networking characteristics has become our technical basis for realizing virtual reality. It includes some techniques such as pattern recognition, remote network, visualization, database, and advanced retrieval. For the application of VR-PDES, in order to enhance the credibility of virtual reality, it must have the ability to evaluate product design with multi-user participation. Therefore, the collaborative environment is very important for this system. It is an extension of interactivity, which refers to multiple users interacting in the same virtual space. The user is aware of the presence of the SIM, allowing users to interact with each other. The collaborative environment can meet the multi-person presence, or multi-person participation mode can meet the comprehensiveness of a system.

The hardware of the VR-PDES consists of a virtual environment generator, input and output devices, and data interfaces.

3. Results and Discussion

3.1. Server Architecture Design and Configuration. The software platform of this paper provides online services for artwork identification and retrieval applications by building a website. The software platform server program selects the model-view-control (MVC) layered architecture that has been widely used at present, in order to reduce the coupling relationship between programs, improve the convenience of system maintenance, and expand the scalability of the system. In the framework selection of system development, this paper selects the current mainstream development frameworks Spring, SpringMVC, and MyBatis (SSM framework). These frameworks can enable the software platform server program of this paper to be completed quickly and integrate the MVC layering idea into the program.

According to the overall scheme and system software planning, the specific functions of the server are expanded in the form of a function tree, as shown in Figure 5. Among them, image understanding and data query are the two core functions of the server. For the sub-module of image understanding, this paper establishes the VGG-16 model to describe the image content of artworks. The deep learning framework used is TensorFlow. This framework not only has advantages in rapid model building, but also has obvious advantages in practical application deployment. Reliability and scalability are advantages, so the server-side image understanding still uses Google's open-source framework TensorFlow for deployment. This module involves a large number of matrix operations and has extremely high requirements for processing capabilities. Therefore, a separate image understanding server is set up to handle this part of the function. In addition, the system deploys a separate main server to carry the remaining functions of data import and data query.

3.2. Database Table Structure Design. Seven relational patterns can be established from the database ER diagram of the artwork recognition retrieval system. These seven data tables corresponding to each relationship mode are scenic spot information in Table 2, all artwork information in scenic spots in Table 3, business information in Table 4, business artwork information in Table 5, e-commerce platform information in Table 6, industry platform electricity platform art information in Table 7, and artwork query information query Table 8. All records involving the description of artwork image content use a binary hash code constructed based on a deep convolutional model.

The scenic spot information table provides the basic information of scenic spots or museums, including name, address, brief introduction, and official website link. The name of each scenic spot or museum must be unique.

Scenic Artwork: The table provides detailed information on the collections or exhibits each scenic spot or museum, and the item names must be unique.

Merchant information: The shop table provides the registration information of merchants in the Wenwan collection industry around the scenic spot on this platform, including name, account password, address, and brief introduction.

The merchant artwork table provides the detailed information of the art and crafts operated by the registered merchants on the platform, in which the merchant number of the item should correspond to the information in Table 4.

The data provided by the platform artwork table are the artwork details of each platform crawled from the crawler.

The platform information table provides information on e-commerce platforms and industry platforms that the system crawls artwork data, including names, brief introductions, and links to the main website.

Retrieval query: The query table provides records of artwork query information, including possible QR codes, binary hash codes generated based on image content, and status flags indicating whether the current retrieval is complete or not.

In addition to the previous information tables that need to be directly stored, the identification retrieval process also requires some intermediate tables to assist in completing the information recording of the retrieval process. The two-dimensional code-based retrieval result table (Table 9) and the binary hash code-based retrieval result table (Table 10) are, respectively, used in the record retrieval process. The processing result of the software platform assists the server and client of the software platform to transmit.

3.3. Image Understanding Module Design and Development. The image understanding module on the server uses the VGG-16 model to perform migration training on the initial artwork dataset in an oversampling manner, then calculates the high-level feature center points of each category, and trains the hash layer of tanh approximate binary quantization encoding. The model parameters and the calculated class center points are saved locally in the image understanding server in the form of configuration files.

The request of the image understanding module can be initiated by the data query module or by the data import module, and the processing result is sent back to the request initiator separately. The specific interfaces are shown in Table 11.

The program that executes image understanding builds the VGG-16 model and binary hash coding model based on the TensorFlow open-source framework. The images in the data import phase are directly read from the local area, and the images in the data query phase are read from the main server in the local area network.

3.4. Data Import Module Design and Development. During data import, the system maintainer batches the prepared artwork graphic data into the database in the form of script calls, and the graphic data uploaded by the merchant are an interface call request initiated by the web client. The text description file during batch import must meet the prespecified format, so as to facilitate batch processing of script programs. The data batch import interface is shown in Table 12, and the web upload text import interface is shown in Table 13. The operation flow of the module is shown in Figure 6.

When querying data, the user uploads the images of the retrieved items using the web client to search for similar artworks and introduction information, or to find similar products on sale and introductions. This module also requires an image understanding module to implement binary hash encoding of image content and image recognition and classification.

The registration and login module here is only provided for the registration of users of merchants near the scenic spot, so that the art products of the merchants can be presented online. Tourists who need to query and retrieve services do not need to register. Before using data import, merchants must complete registration and login to have upload permission. The merchant initiates a registration request, the client prompts to enter the relevant information and confirms the server calls the registration verification module to confirm the validity of the registration information, the account can be assigned if the rules are met, and

Table 2: Definition of scenic spot information table.

Id name	Id class	Id constraints	Id description
Scenery id	Int (8)	Primary key	Number
Scenery name	varchar (45)	Not null, unique	Name
Scenery located at	varchar (45)	Not null	Address
Scenery description	varchar (200)	Not null	Introduction
Scenery url	varchar (100)	Not null	Link address

Table 3: Scenic artwork table definition.

Id name	Id class	Id constraints	Id description
Artwork id	int	Primary key	Number
Artwork name	varchar (45)	Not null, unique	Name
Scenery id	int (8)	Not null, foreign key	Scenery spots id
Artwork bcode	char (6)	Not null	Binary hash code of the item image
Artwork description	varchar (200)	Not null	Introduction
Artwork origin located at	varchar (45)	Not null	Address
Artwork origin time	varchar (45)	Not null	Age
Artwork url	varchar (100)	Not null	Link address
Artwork picFile	varchar (100)	Not null	Figure address
Artwork QR code	varchar (100)	—	QR code

Table 4: Definition of shop information.

Id name	Id class	Id constraints	Id description
Shop id	int (8)	Primary key	Number
Shop name	varchar (45)	Not null, unique	Name
Shop pawd	varchar (16)	Not null	Account password
Shop located at	varchar (45)	Not null	Address
Shop description	varchar (200)	Not null	Introduction

Table 5: Definition of merchant artwork.

Id name	Id class	Id constraints	Id description
Artwork id	int	Primary key	Number
Artwork name	varchar (45)	Not null	Name
Shop id	int (8)	Not null, foreign key	Shop id
Artwork bcode	char (6)	Not null	Binary hash code of the item image
Artwork description	varchar (200)	Not null	Introduction
Artwork price	decimal (6,2)	Not null	Price
Artwork picFile	varchar (100)	Not null	Figure address

Table 6: Industry platform e-commerce platform artwork table definition.

Id name	Id class	Id constraints	Id description
Artwork id	int	Primary key	Number
Artwork name	varchar (45)	Not null, unique	Name
Platform id	int (8)	Not null, foreign key	Shop id
Artwork bcode	char (6)	Not null	Binary hash code of the item image
Artwork description	varchar (200)	Not null	Introduction
Artwork price	decimal (6,2)	Not null	Price
Artwork url	varchar (100)	Not null	Link address
Artwork picFile	varchar (100)	Not null	Figure address

the merchant registration information is entered into the database and returns a successful registration response; otherwise, it returns a failure response. When the merchant logs in again, the server calls the merchant login module to verify the validity of the account and password, and responds to the information entry page if it is legal; otherwise,

Table 7: Definition of platform information.

Id name	Id class	Id constraints	Id description
Platform id	int (8)	Primary key	Number
Platform name	varchar (45)	Not null, unique	Name
Platform description	varchar (200)	Not null	Introduction
Platform url	varchar (100)	Not null	Link address

Table 8: Lookup table definition.

Id name	Id class	Id constraints	Id description
Query id	int	Primary key	Number
Query bcode	char (6)	Not null	Query the binary hash code of an image
Query QR code	varchar (100)	—	QR
Query finished	tinyint	Not null	Retrieval completion status flag

Table 9: Definition of search result table based on QR code.

Id name	Id class	Id constraints	Id description
Artwork id	int	Primary key	Number
Resource	tinyint	Primary key	Search result source indication (0/all 1/scenery spot 2/shop 3/website)
Query QR code	varchar (100)	—	QR
Query id	int	Not null	Inquiry number

Table 10: Retrieval result table definition based on binary hash code.

Id name	Id class	Id constraints	Id description
Artwork id	int	Primary key	Number
Resource	tinyint	Primary key	Search result source indication (0/all 1/scenery spot 2/shop 3/website)
Distance	tinyint	Not null	Hamming distance
Query id	int	Not null	Inquiry number
Artwork picFile	varchar (100)	Not null	Figure address
Artwork description	varchar (200)	Not null	Introduction

the login data are reset to empty and a corresponding error is displayed.

3.5. Web Client Development. The paper uses JSP and JavaScript technology in designing the web client. In a broad sense, JSP is a dynamic web page technology, which converts dynamic web pages into web pages through back-end Java program processing and transmits them to browser clients. The view layer of the MVC layered idea is mainly embodied as an intuitive and operable web interface in the software platform of this paper.

Based on HTML5, this paper combines the scripting language JavaScript to add dynamic features to the web client, such as event response, so as to make the page interaction effect better and optimized. In addition, the appearance design of the page adopts CSS3, which provides pixel-level control for web page display, such as setting web page fonts and colors.

When designing the view layer, in order to separate the code unrelated to the business logic from the interface, to decouple the view layer and the controller, and to facilitate later maintenance and secondary development, this paper mainly uses jquery when developing the web client. Ajax technology and JSON data format are used for front-end and back-end communication. After the JSON data protocol between the front-end and the back-end is formulated, data are requested from the server through JavaScript, and the interface is re-rendered after getting the response to update the data. The interaction between JavaScript and the back-end usually adopts the asynchronous communication mode; that is, the browser will not enter the response wait after sending a request to the server, but will continue to execute the subsequent code. When the server returns the data, the browser will execute the message response function to complete the response action.

The interface of the web client includes four pages: photo query, retrieval display, business registration and login, and information entry.

The photo query page is the main page that provides users with query operations. This page includes query image upload function, photo upload image function, and QR code scan

Table 11: Image understanding module request interface definition.

Item	Description
Method name	ImageParse
Call method	TCP communication trigger call in LAN
Request parameters	Parameter 1: String queryImage, query image path identifier Parameter 2: Int queryType, query type identifier, 1 QR code, 0 photo image Parameter three: Int queryFrom, the request originating source identifier, 1 for query, 0 for import
Return parameter	String, which needs to be parsed again and restored to binary hash code/QR code/identification category

Table 12: Data batch import module interface definition.

Item	Description
Method name	ArtworkImport
Call method	Console script call, TCP communication trigger call in LAN
Request parameters	Parameter 1: String artwork_source, import image path identifier Parameter 2: String artwork_descript, the path identifier of the text description file Parameter three: Int artwork_type, import type ID (1 QR code, 0 images) Parameter 4: Int artwork_from, import source ID
Return parameter	Boolean, import success flag

Table 13: Merchant upload text import module interface definition.

Item	Description
Inquiry address	https://223.3.86.94:8808/shop_artwork/upload
Inquiry method	POST
Response format	JSON
Inquiry parameter	Parameter 1: String artwork_name, the name of the imported artwork image Parameter 2: Int shop_id, merchant ID Parameter three: String artwork_description, brief description Parameter four: Float artwork_price, price information
Responsive parameter	—
Responsive code	00, 201, 400, 401, 403, 404, 408, 500

upload function. The retrieval display page is the identification retrieval result display page after the user clicks the search, and the page includes the identification result and the retrieval item result. The search item contains the result image and the corresponding text content. The entire display page also provides filtering and filtering operations for search results, so that users can view different search sets separately. The realization function of the merchant registration login page is to register the scenic spot merchants and upload artwork images. The page prompts you to enter relevant information and confirm. The server calls the registration verification module to check the legitimacy of the user information. If it is legal, the account is created and the image and text information is allowed to upload. The client information entry page is a page that provides the merchant with the operation of uploading artwork images and texts after the merchant logs in. This page includes image file upload operations and text-related information input operations. After the merchant completes the entry of the information on the page, the server calls the data import module to check the validity of the information imported by the merchant. If it is legal, the artwork data uploaded by the merchant are entered into the system database, and the image understanding module is called to produce the image content binary of the image of the item. Hash code, these information will be entered into the system database together, and the entry success response will be returned; otherwise, the entry failure response will be returned.

3.6. System Frame Model Diagram for VR-PDES. Based on the previous content analysis and research, this study proposes a product design evaluation system based on virtual reality technology—VR-PDES. VR-PDES frame is shown in Figure 7. The environment for evaluation work is jointly constructed by the virtual reality system and the product modeling design system driven by design intent. The system is constructed by the design evaluation information data obtained in the virtual reality system. As a user experience, the virtual environment provides a user environment with a strong sense of immersion, and the tracking and evolution of user experience information are obtained through the interaction between the user and the virtual reality system.

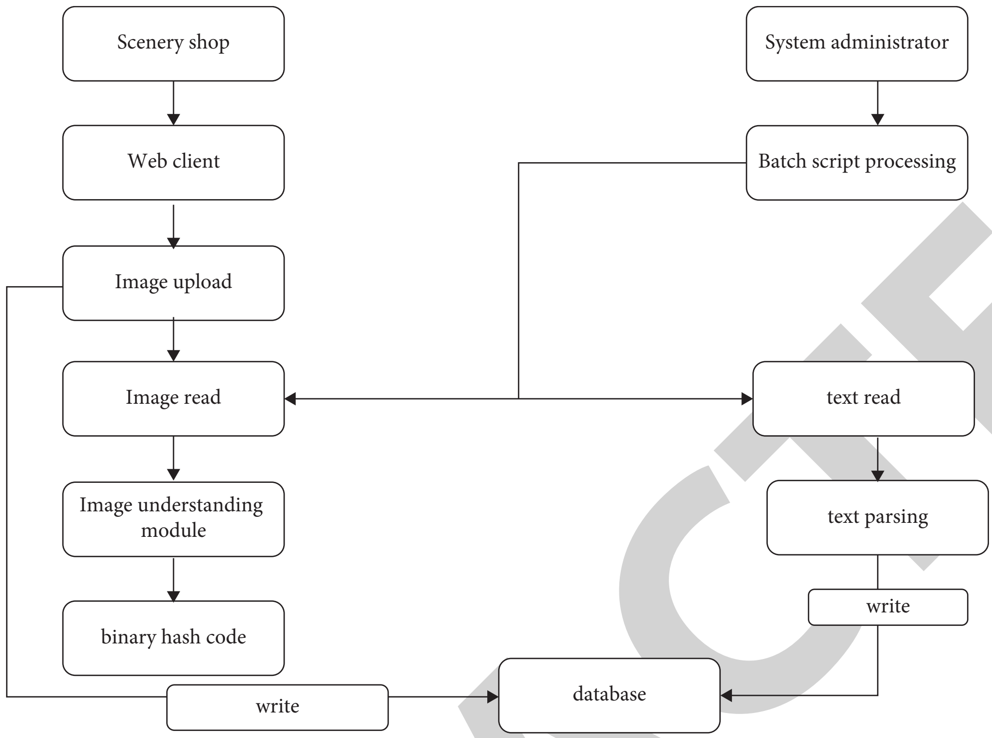


Figure 6: Data import operation process.

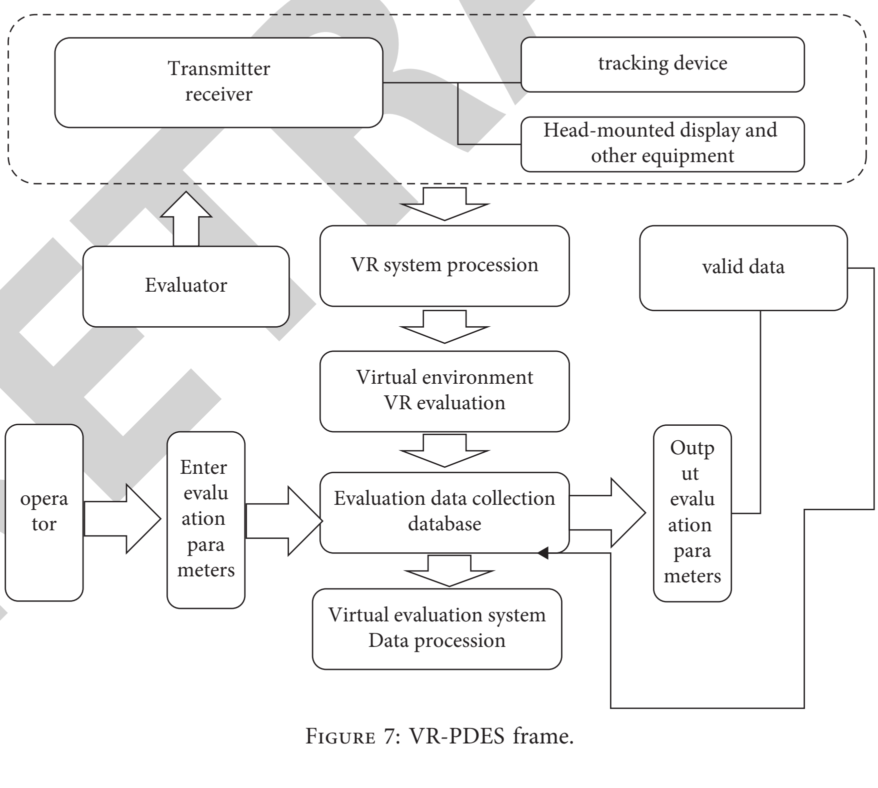


Figure 7: VR-PDES frame.

4. Conclusions

In our study, we first studied the image-based artwork recognition and retrieval solution, discussed the system function and architecture in detail, analyzed and compared the system database, and completed the data requirement discussion and conceptual design, and then grabbed artwork-related graphic data from multiple platforms through crawler programs, summarized all the data to complete a large-scale artwork dataset, and analyzed the imbalanced state of image dataset categories; moreover, the SSM framework is used to complete the development of all

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9826374, 1 page
https://doi.org/10.1155/2023/9826374

Retraction

Retracted: Practice and Exploration of Teaching Mental Health Education for College Students Based on Data Mining Algorithm

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Bai and C. Tang, "Practice and Exploration of Teaching Mental Health Education for College Students Based on Data Mining Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 8936152, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8936152, 9 pages
https://doi.org/10.1155/2022/8936152

Research Article

Practice and Exploration of Teaching Mental Health Education for College Students Based on Data Mining Algorithm

Lu Bai[1] and Chunyan Tang[2]

[1]*Xi'an Siyuan University, Xi'an, Shaanxi 710038, China*
[2]*Jiangxi Health Vocational College, Nanchang Medical College, Nanchang 330052, China*

Correspondence should be addressed to Chunyan Tang; tangchunyan@ncmc.edu.cn

Received 30 March 2022; Revised 24 April 2022; Accepted 6 May 2022; Published 17 June 2022

Academic Editor: Chin-Ling Chen

Based on the data mining algorithm, this paper combines the background of the requirements for cultivating talents in the new era, the psychological characteristics, and learning features of contemporary college students and the realistic needs of the reform of the teaching of mental health education courses in colleges and universities. By analysing the characteristics, psychology, and causes of crime of contemporary university students, the study will explore the regular experience of guiding and preventing university students from committing crimes at the level of mental health education in universities, which will not only help to reduce the crime rate of university students but also extend to the prevention of crimes of other people in society through the work of mental health education in general, so as to reduce the crime rate in society and create a harmonious society. In this paper,the decision tree C4.5 algorithm and the association rule apriori algorithm are used to analyze the mental health data of college students, which can improve the teaching quality of college mental health education courses and enhance the psychological quality of college students. It is of great significance to improve the teaching quality of mental health education courses in colleges and universities and to enhance and improve the psychological quality and health literacy of college students.

1. Introduction

University students are the valuable human resources of the country, the hope of the nation, and the future of the motherland. How they think and behave has a great deal to do with the future and destiny of the country. It is of great and far-reaching strategic significance to ensure that the cause of socialism with Chinese characteristics flourishes and is succeeded by people [1].

Compared with other people of the same age, college students should have a higher moral quality. Although the overall quality of contemporary college students has generally improved, there are still individual college students with low moral quality, especially the growth of the crime rate of college students is even more worrying [2]. According to the relevant survey, "in 1965, the crime of college students accounted for 1% of the criminal crimes of the whole society," during the "Cultural Revolution," the crime of college students accounted for 2.5% of the whole criminal crimes, while in recent years, the crime of college students accounted for 17% of the whole social criminal crimes [3]. "Since the beginning of the new century, reports on crimes committed by university students have come into the limelight even more frequently and have become a matter of concern for society at large." For example, the major vicious crime cases such as the case of Yao Jiaxin in 2010, the stabbing case of a mother at the airport of a foreign student in 2011, the assassination case of the vice president of Jiangsu University of Science and Technology in 2012, and the murder case of Li Starr at the Communication University of China in 2015 have all caused people to pay attention to and think about the crimes committed by university students [4]. As a group with a special status in society, university students have a bright future and a bright future. The state nurtures them and the schools teach them knowledge and skills in the hope that they can realize their value in life, support the future of the motherland, and contribute to society. This is a real warning to today's university education [5].

The role of colleges and universities is to teach and educate people, but some colleges and universities put more effort on academic education for college students, but to a certain extent neglect the legal education of college students [6]. A highly educated criminal may be more sophisticated, technical, and intelligent than an ordinary criminal, which undoubtedly creates more danger for the victim and makes the detection of the case more difficult. In the face of the rising crime rate among university students, universities need to rethink their current education [6–8].

There are still some weak links in the implementation process of mental health education in colleges and universities, and the degree of attention paid to it still needs to be improved [9]. Exploring how to effectively prevent crimes among college students and effectively improve the effectiveness of mental health education in colleges and universities will help colleges and universities improve the level of mental health education and provide guarantee for improving students' moral quality, shaping a good campus climate, cultivating a good learning atmosphere, and building a harmonious campus [10].

Therefore, as an important research content of mental health education, it is of great theoretical and practical significance to strengthen the research on countermeasures for the crime problem of college students. Therefore, as an important research content of mental health education, it is of great theoretical and practical significance to strengthen the study of countermeasures for college students' crime problems.

2. Review of the Literature

Since the end of the last century, there have been many scholars who have begun to study the crime problem of college students, most of them are in the field of criminology and sociology, etc., to explore and study the characteristics, types, causes, and preventive measures of college students' crimes. [11] According to the different motives of delinquency and crime, the delinquency and crime of college students are divided into four types: property type, sexual sin type, revenge type, and political type. [12] According to the standard of criminology, the crimes of college students are divided into three categories: property crimes, violent crimes, and sexual crimes.

Most scholars believe that the characteristics of current crimes committed by university students are increasing in number, diverse types, highly intelligent and brutal crime methods, and the expansion of the subjects involved in crimes [13] argues that "college students' crimes present characteristics such as the young age of the offender, impulsiveness, extreme ego and individualism, concentration of crimes in property and violence type, clear purpose of committing crimes, and simple motives for committing crimes." [14] summarises the characteristics of crime among university students as: diversity, passion, high intelligence, and cruelty. [15] argues that, "College students are socially inclined to commit crimes. Higher education institutions are part of the society, constantly subject to the impact and influence of the general social environment, and under the temptation of money, many college students collude with the unruly elements in the society and engage in illegal and criminal interaction. For example, some university students and some criminals in society have joined forces to engage in fraud. This characteristic of socialization of crime will become more and more obvious and will become the focus of crime prevention among college students."

Mental health education can play a very obvious role in the prevention of crime among university students. The role of mental health education is identified in [16] as, "The object of mental health education is people, and the goal is to promote the overall development of people. The functions of mental health education are mainly orientation function, guarantee function, nurturing function and development function. Preventing college students from committing crimes is a direct requirement of the guiding function of mental health education's behavioral norms" [17]. "The formation and development of people's mental health and moral and legal qualities are based on psychological activities, and in order to make mental health education achieve effective internalization and externalization from thoughts to behaviors, mental health education needs to pay special attention to the role played by the process of human psychological development in bringing into play the timeliness of mental health education." [18] Strengthening mental health education for college students can help them establish ideals and beliefs, guide them to conform to the requirements of our socialism morally and legally, effectively prevent crime, and become useful talents of socialism.

3. The Feasibility of Mental Health Education to Address Crime Prevention among University Students

Mental health education is a compulsory course to systematically educate college students on ideology and morality and general knowledge of social sciences, and it has a guiding function to enhance the quality and cultivation of college students [19]. Through the role of mental health education, the prevention of crime as the purpose of mental health education for college students can achieve good results.

3.1. Playing a Guiding Function to Establish the Correct Values of University Students. College students are highly malleable and their values are still in the process of formation. In a society with diverse ideas and cultures, information of good and bad quality can easily make their values deviate. Mental health education can lead university students' deviant values to the mainstream by guiding, restraining and inspiring them. It enables university students to set up requirements that are in line with those of the party and the state. It also sets an example through behavioural guidance, so that everyone can follow suit in terms of spirit, psychology, personality, and behaviour to achieve self-restraint and recognition of social norms among university students.

3.2. Performing a Normative Function to Regulate the Behaviour of University Students. Mental health education in higher education can enhance the legal awareness and moral quality of university students. It enables university students to deal with social conflicts, emotional entanglements, and psychological misunderstandings in a coordinated manner according to socially accepted rules, without being emotional or blindly impulsive. Awareness of rules is the minimum quality and ability of social people. The cultivation of awareness of rules and responsibility in mental health education in colleges and universities becomes an important guarantee for the standard of behaviour of college students.

3.3. To Play a Coordinating Function to Adjust Interpersonal Relationships among University Students. Social interaction is a basic way of life for university students, so play the coordinating role of mental health education, smooth the channels of interpersonal interaction among university students, improve their social interaction ability and quality, resolve conflicts, coordinate interpersonal relationships, and avoid many unnecessary disputes.

4. Practical Pathways for Teaching Mental Health Education

4.1. Creating Deep Learning Classrooms with Pedagogical Intelligence. By selecting real-life cases that contain psychological principles, have educational significance and are enlightening and research-oriented for students' minds, teachers promote students' active exploration of mental health cases in relation to their own learning and life, and obtain more information that can be applied, analysed, and evaluated to effectively solve the mental health problems they encounter. Again, teachers should engage in moderate self-expression in the classroom, combining mental health knowledge with personal life experiences, discussing with students how their own professional, spiritual, or moral and emotional lives are intertwined (and sometimes conflicted) and how they have faced and resolved such problems, providing valuable and learnable modes of thinking for students' growth, which will better facilitate students' deeper learning, enabling students to connect mental health knowledge to their own learning and lives, to build personalised meanings of what they have learned, to apply what they have learned in the classroom to their real-life learning and lives, and to promote real-life problem solving and new creativity.

4.2. Student-Centred, Identifying the Learner as the Subject. The concept of "student-centred" and focusing on students' deeper learning is the foundation of a high-quality undergraduate education. First, teachers should analyse the students' situation and get to know them comprehensively, not only by analysing their existing mental health knowledge but also by understanding the learning, life, and psychological characteristics of students in the new era, and by tapping into their personal learning and life experiences to bring in material from their own lives for classroom learning, so that they can participate more deeply in classroom learning. Second, as students are the subjects of cognition and active constructors of knowledge meaning, teachers should take the initiative to explore and study how to better utilise the features of the online teaching mode to carry out mental health education teaching activities in the context of digital learning, with students' learning at the centre, to enhance students' interest in learning, promote the role of endogenous factors in students' learning, cultivate students' initiative and creativity, and guide them to deep learning.

Teachers can therefore assign exploratory writing to students as a classroom assignment in mental health education, which can guide students in deeper learning while also providing indicators for process- and development-based assessment of achievement.

4.3. Create Effective Learning Contexts to Promote Student Engagement. Contextual cognitive theory suggests that knowledge is generated as learners interact with social and physical contexts. Creating effective learning contexts guided by contextual cognitive theory can better support and facilitate the occurrence of deep learning. First, it is important to create a safe classroom learning context. As the content of mental health education courses will involve students' knowledge, inner experiences, and feelings about themselves and others, there is a need to create a trustworthy and safe learning context where students have the opportunity to self-express in an appropriate way, encourage them to openly express their true experiences and feelings in class, and be willing to discuss them with peers or teachers around them to promote personal reflection and discover the connections between old and new knowledge, or recognise the relevance of what they have learned to their learning and lives, and deepen their understanding, transfer, and application of mental health knowledge. Second, it is important to create a participatory, experiential, and interactive learning context in the classroom so that students learn knowledge and skills about mental health in a lively and active teaching and learning environment. An experiential and interactive classroom enables students to integrate their own active understanding and feelings into a state of "moving," generating emotion and meaning, learning through conflict and joy, forming collisions between their own and others' minds in performance and discussion, experiencing the rehearsal or realisation of various psychological needs, allowing emotional learning to take place in the classroom, and hitting students directly in the heart and soul. This allows emotional learning to take place in the classroom, to rethink the self, to rebuild the true self, to enhance problem-solving skills, and to promote the effective application of classroom teaching in real life. Once again, a culture of integration and symbiosis is a guarantee for the design of in-depth learning. In addition to psychology, the theoretical foundation of university students' mental health education courses must have a high degree of cultural confidence, i.e., using the essence of Chinese culture to construct the Chinese heart, Chinese feelings, and the character and charm of Chinese people among university students. We should explore the unique thinking characteristics, behavioural habits, and value patterns of Chinese students imbued with Chinese culture and build a mental health education curriculum for university students with Chinese cultural characteristics.

5. Data Mining Applied to Student Mental Health Research

After years of development, data mining techniques have been used in a wide range of fields with good results that are unmatched by other technical methods. Data mining techniques are used to uncover the hidden laws and values in data to solve specific problems, and people can use computer applications to perform the same functions when they do not have the technology to do so. Therefore, this paper proposes to introduce data mining technology into the study of mental health data, mining and analysing the causes of students' mental health problems, providing a scientific basis for early prevention and early intervention to control the emergence of psychological crises on campus.

5.1. ID3 Algorithm. Proposed by Quinlan in 1986, ID3 is a well-known algorithm in machine learning and is the first and most influential decision tree method proposed in the world [9]. In the ID3 algorithm, the most important feature is the selection of attributes. The attribute with the greatest gain will be used as the root node of the decision tree, and it is necessary to define the gain using entropy, one of the concepts of information theory. The information gain of the prepartition entropy and postpartition entropy is calculated as a criterion for judging the information content of an attribute [10].

Suppose P is the set of 5 data samples. Then the category attributes that have M different values: $C_i \quad (i = 1, 2, \ldots, m)$. Let S_i be the number of samples in class C_i. For this given sample, its total information entropy is

$$I(s_1, s_2, \ldots, s_m) = -\overset{\infty m}{a}_{i=1} P_i \log_2 (P_i), \tag{1}$$

where P_i is the probability of any sample belonging to C_i, which we can usually express as s_i/s.

Assume that attribute A has a different value $\{a_1, a_2, \ldots, a_k\}$, that the set S is divided into no subsets $\{s_1, s_2, \ldots, s_k\}$ by attribute A, and that the samples in the set S for which attribute A takes the value a_j are included in subset s_j. If the test attribute is determined to be attribute A, then the new leaf node generated from this set node is the subset described above. Let s_{ij} be the number of samples in subset s_j with category C_i, and the information entropy value of the sample obtained by partitioning is

$$E(A) = {}^{k}_{j=1} \frac{s_{1j} + s_{2j} + \cdots + s_{mj}}{s} I\left(s_{1j}, s_{2j}, \ldots, s_{mj}\right), \tag{2}$$

where $P_{ij} = s_{ij}/s_{1j} + s_{2j} + \cdots + s_{mj}$ is the probability of a sample with category C_i in subset s_j.

Finally, the information gain (Gain) of the sample set S is obtained by dividing:

$$\text{Gain}(A) = I(s_1, s_2, \ldots, s_m) - E(A). \tag{3}$$

Show that if the value of information entropy $E(4)$ becomes smaller, the value of information gain Gain(A) becomes larger. Then the uncertainty of the test attribute A on the classification becomes small.

5.2. C4.5 Algorithms. Original ID3 algorithm mainly made the following adjustment changes.

Split Information for Attribute A:

$$\text{Split}\, I(A) = -\sum_{j=1}^{m} \frac{|S_j|}{|S|} \times \log_2 \frac{|S_j|}{|S|}. \tag{4}$$

In the above equation, attribute A splits the training data set S into m sub data sets, and the number of samples in the jth sub data set is denoted as $|\mathbf{S}_j|$, while $|\mathbf{S}|$ is the total number of samples in the data set before the split.

After splitting by attribute A, the value of the information gain rate of the sample set is

$$\text{Gain Ratio}(A) = \frac{\text{Gain}(A)}{\text{Split}(A)}. \tag{5}$$

When the decision tree is constructed and generated by the C4.5 algorithm, the splitting attribute of the current node is determined by the attribute with the maximum information gain rate. The information gain rate of the attribute is calculated to become progressively smaller, and the attribute with the relatively large information gain rate is determined as the splitting attribute when it is generated later.

The decision tree-based C4.5 algorithm operates as follows.

(1) First, the information gain rate of each split attribute in the training sample set is obtained by calculation.
(2) The root node of the decision tree is determined by the splitting attribute with the largest information gain rate, and the data set is split into corresponding sub data sets according to the number of values taken.
(3) Step 1 and 2 are performed recursively in the sub data set in turn.

5.3. Association Rule Algorithm. Association rules reflect the interdependence and correlation between one thing and other things and are used to explore the correlation between valuable data items from a large variety of data.

(1) Representation of association rules

The probability of item set A and item set B occurring together is called the support of an association rule, also known as relative support:

$$\text{Support}(A \Rightarrow B) = P(A \cap B). \tag{6}$$

When item set A occurs, the probability that item set B also occurs is the confidence level of the association rule:

$$\text{Confidence}\,(A\Rightarrow B) = P\,(B, A). \tag{7}$$

(2) Minimum support and minimum confidence

The threshold used to define a measure of support is often referred to as the minimum support, which represents the minimum statistical significance criterion for an item set. The threshold used to define a measure of confidence is often referred to as the minimum confidence level, which represents the minimum standard of reliability of an association rule. Rules that meet both the minimum support threshold and the minimum confidence threshold are often referred to as strong rules.

(3) Item sets

A general frequent k-item set can be written as L_k

(4) Support counts

If the support counts of the item set are known, the support and confidence of rule $A\Rightarrow B$ can be deduced from the support counts of all transactions, item set A and the support counts of the item set:

$$\begin{aligned}\text{Support}\,(A\Rightarrow B) &= \frac{A, B\text{Number of simultaneous transactions}}{\text{Number of all transactions}} = \frac{\text{Support\_count}\,(A\cap B)}{\text{Support\_Count}\,(A)},\\ \text{Confidence}\,(A\Rightarrow B) &= P\,(B \mid A) = \frac{\text{Support}\,(A\cap B)}{\text{Support}\,(A)} = \frac{\text{Support\_count}\,(A\cap B)}{\text{Support\_count}\,(A)}.\end{aligned} \tag{8}$$

As soon as the number of all transactions, iteration 5 and the support counts of the three item sets are obtained, the corresponding association rules $J5$ and $5d$ are generated, which ultimately determine whether they are strong association rules or not.

The basic idea of the Apriori algorithm is to retrieve all frequent itemsets from the transaction database and find the maximum frequent itemset and a predefined minimum trust construct to generate a strong association rule.

(1) Properties of Apriori

All nonempty subsets of a frequent itemset are also frequent itemsets. According to this property, we can conclude that if we add transaction A to the set of items that are not frequent, the newly generated itemset $I\cap A$ cannot be a frequent itemset.

(2) The process of implementing the Apriori algorithm:

(a) By finding all frequent itemsets, where the support cannot be less than the minimum support threshold set, and by connecting and pruning the two steps in the process of finding, the largest frequent itemset L can be obtained after several cycles of operation.

5.4. Mind Health Data Mining Business Processes. The SCL-90 was used to collect data on mental health problems among university students. The SCL-90 is a world-renowned measure of mental health that is widely used in the diagnosis of mental illness and detection of mental disorders, analysing the extent of a patient's mental health status in 10 dimensions. The scale has 90 items SCL-90 rated on a 5-point scale (or 4-point scale) and contains a wide range of psychiatric symptomatology, including thinking and feeling, eating and sleeping, interpersonal relationships, behavioural habits, etc., from 10 factors (obsessive-compulsive symptoms, depression, somatisation, hostility, anxiety, interpersonal sensitivity, phobia, psychoticism, paranoia, and other symptoms). Psychoticism, paranoia, and others) reflecting 10 aspects of psychological symptomatology. The results are a reflection of the level of psychological symptoms over time and have the ability to distinguish between people with psychological health and those with psychological disorders. It is suitable for use in testing the presence or absence of symptoms of mental health problems in a population. It is generally used in hospitals for the detection of clinical psychological symptoms and can also be used for the screening of new university entrants for mental health.

Due to the many iterations in data mining, the mining tools used in various fields are different in the widely used data mining techniques. The selection of appropriate data mining algorithms and data mining tools to guide computer analysis to obtain valuable knowledge information is always an important part of data mining efforts. By using the classification rule C4.5 algorithm and the association rule Apriori algorithm, the data mining operation on college students' mental health data is carried out through the wood paper.

The data mining process of college students' mental health assessment data is shown in Figure 1.

The database system of a health vocational college used SQL Server 2008 to store and manage students' personal information, including student number, ID card number, name, gender, ethnicity, date of birth, place of origin, household registration, department, major, only child or not, and contact telephone number [20, 21]. The student's personal psychological assessment form is generated by the psychological assessment system and reflects the student's tendency to have personal psychological problems through the relevant psychological dimensions. It contains the student's student number, name, topic 1 . . . And 10 psychological dimensions: obsessive-compulsive symptoms, depression, somatisation, hostility, anxiety, relationship sensitivity, phobia, psychoticism, paranoia, and other fields, as shown in Table 1.

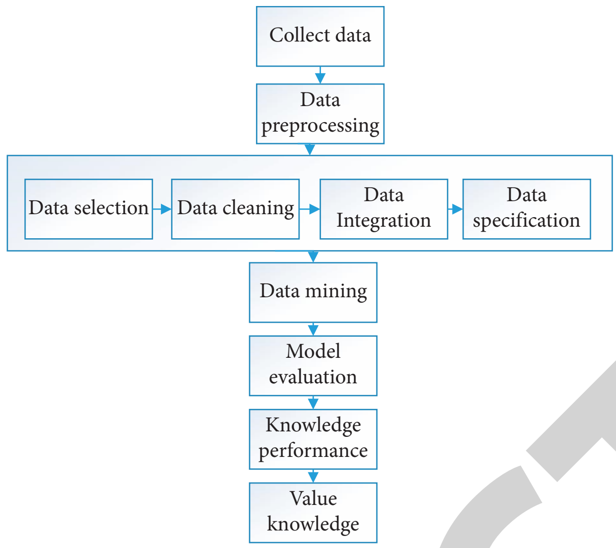


FIGURE 1: Flow chart for data mining of mental health assessment for university students.

TABLE 1: Selected data content after data integration.

Gender	Registered residence	Family status	Faculty	Somatization	Obsessive compulsive symptoms	Depressed	Anxious	Psychotic	Other
Female	Town	Yes	Non needy households	1.33	1.5	1.31	1.4	1.1	1.29
Female	Town	Yes	Non needy households	1.17	2.1	1.92	1.6	1.5	1.71
Female	Countryside	No	Difficult family	1.92	2.3	2.15	2.2	2	1.86
Female	Countryside	No	Non needy households	1.92	2.1	2	2.1	1.6	1.57
Female	Countryside	No	Non needy households	1.33	2.1	1.15	1.3	1.2	1.43
Female	Town	Yes	Non needy households	1.5	2.3	2.77	2.4	2.4	1.86
Female	Countryside	No	Non needy households	1.08	1.6	1.23	1.2	1.1	1
Female	Countryside	No	Non needy households	2.33	2.4	2.23	1.7	1.3	1.71
Female	Countryside	No	Non needy households	1.5	1.7	2.31	1.3	1.4	1.29
Female	Countryside	No	Non needy households	1.17	1.2	1.31	1.1	1	1.29
Male	Town	Yes	Non needy households	2.08	2.7	1.92	2.6	2.2	2.14
Female	Countryside	No	Difficult family	2.58	2.9	2.46	2.4	2.2	2.43
Female	Town	Yes	Non needy households	1	1.3	1	1	1	1.14
Female	Countryside	No	Difficult family	1.17	1.9	1.54	1.6	1.5	1.14
Male	Countryside	Yes	Non needy households	1	1	1	1.1	1.1	1

Data integration is the process of integrating records from multiple related data collections into a new data warehouse based on the content of the mining objectives. The data used in the thesis are mainly from the Basic Student Information Form and the SCL-90 Mental Health Assessment Form. The two tables are linked through the association fields, and the data set determined by the "data selection" process is used to generate a new mental health assessment form for students, as shown in Table 2.

6. Effectiveness of Mental Health Education

From the analysis of the causes of crimes committed by university students, it can be seen that many university students commit crimes because they do not have a firm political stance and a high moral quality. And the cultivation of socialist ideological character and the improvement of the political and ideological awareness and moral level of all people is one of the fundamental goals of mental health

Table 2: Mental health assessment scale for university students (partial data).

XB	HK	DSZN	JTZK	BX	QTH	QPZZ	RJGX	YY	JL	DD	KB	PZ	JSPX	QT
XB2	HK1	DS1	JT2	BX1	QTH2	QP2	RJ2	Y1	JL2	D2	KB2	PZ2	JS2	QT2
XB2	HK1	DS1	JT2	BX1	QTH2	QP1	RJ2	Y1	JL2	D2	KB2	PZ2	JS2	QT2
XB2	HK2	DS2	JT1	BX1	QTH2	QP1	RJ1	Y2	JL1	D1	KB1	PZ1	JS2	QT2
XB2	HK2	DS2	JT1	BX1	QTH2	QP1	RJ1	Y2	JL1	D1	KB1	PZ2	JS2	QT2
XB2	HK1	DS2	JT2	BX1	QTH2	QP1	RJ2	Y1	JT2	D2	KB2	PZ2	JS2	QT2
XB2	HK2	DS1	JT2	BX1	QTH2	QP1	RJ1	Y2	JT1	D1	KB1	PZ1	JS1	QT2
XB2	HK1	DS2	JT2	BX1	QTH2	QP2	RJ2	Y1	JT2	D2	KB2	PZ2	JS2	QT2

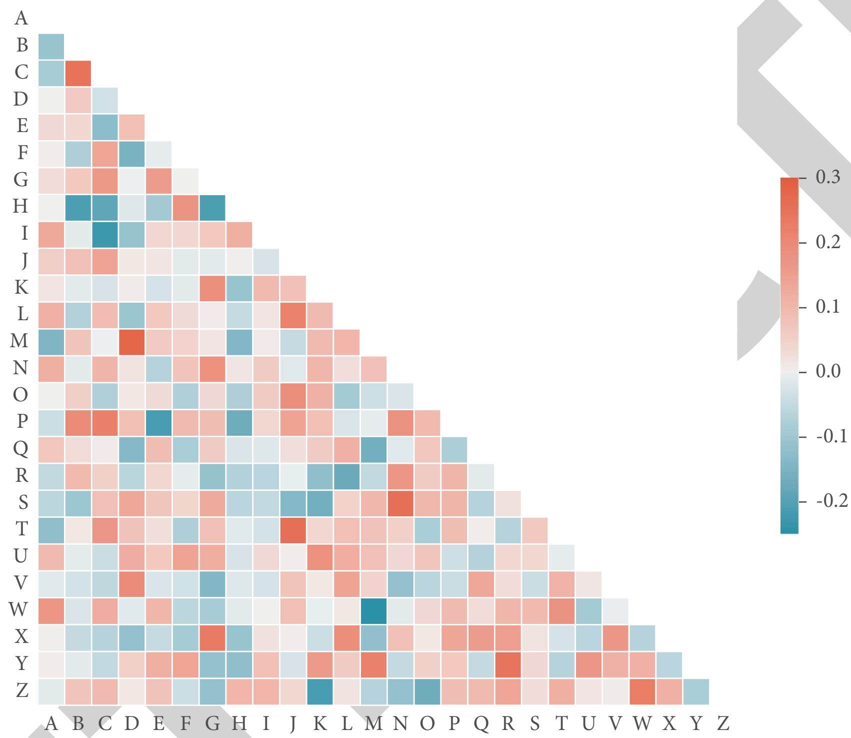


Figure 2: Heart health energy map.

education. As shown in Figure 2, through the energy diagram of the psychological changes of university students after the data analysis in this paper, we can know that mental health education can guide university students to master the scientific worldview and methodology, so that they have the initial ability to draw the line between materialism and materialism, science and superstition, civilization and ignorance, and resist materialism, feudal superstition, and all kinds of pseudoscience [21, 22].

As shown in Figure 3, psychological problems are one of the major causes of crime among university students, and many university students commit criminal acts not because of moral quality but because of psychological problems, so it is very important to ensure psychological health, both for society and for individual university students. Mental health education can educate and guide university students, through persuasion and motivation, strengthen psychological counselling for different individual requirements, focus on guiding university students to form positive emotions, moderate emotions, harmonious interpersonal relationships, good personality qualities

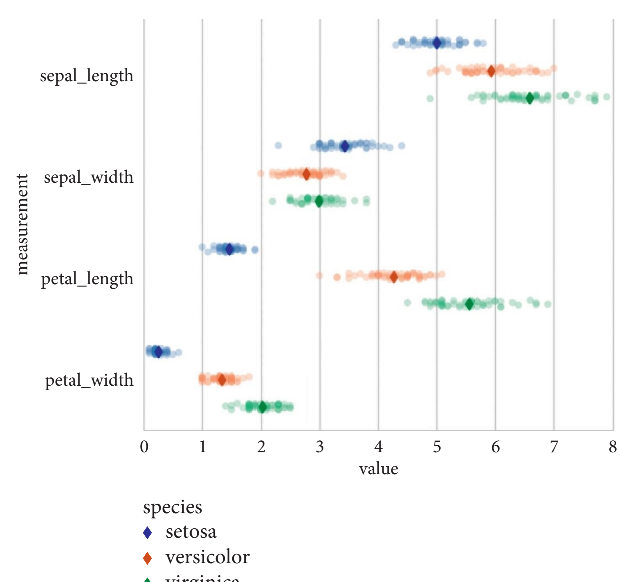


Figure 3: Propensity to offend after psychoeducation.

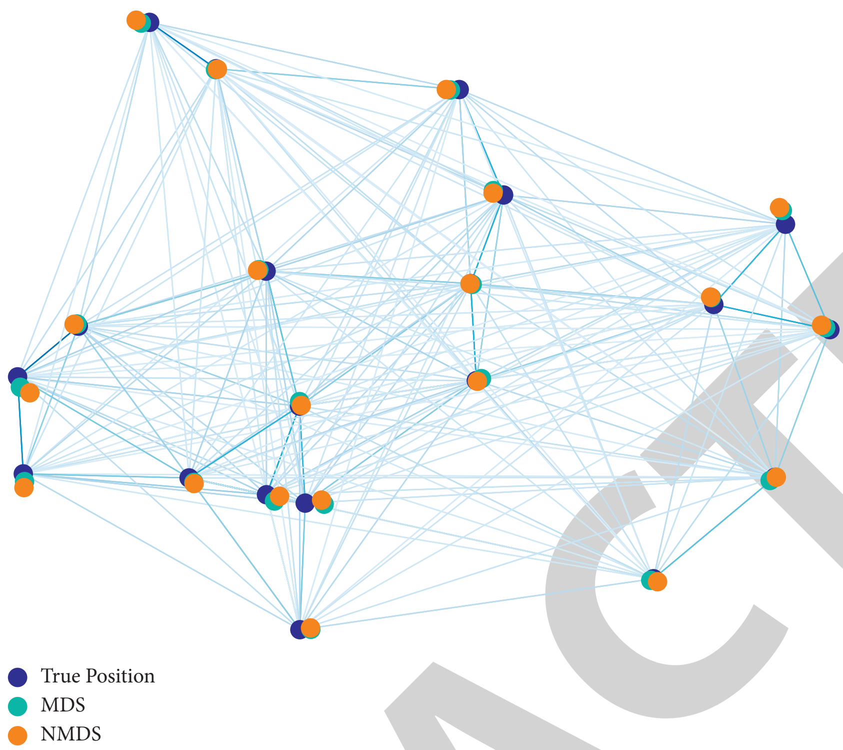


Figure 4: Different factors associated with university students.

and a strong will, and help them develop healthy psychological qualities, thus eliminating the situation of university students committing crimes due to psychological problems.

College students have strong self-esteem and competitive spirit during adolescence, and because they come from different regions and families, they have obvious differences in their lifestyles, ideologies, and moral standards, and conflicts and contradictions can easily arise between them. The association of different factors is shown in Figure 4, where the study shows that 30% of vocational school students have some psychological barriers to communication, lack self-confidence, and are overly cautious and timid, leading to severe psychological depression. In their interactions with teachers, peers, and family members, not wanting to open up about their ideas, their ability to adapt, academic problems, interpersonal relationships, etc., students usually show loneliness, suspicion, jealousy, reluctance, or fear of communicating with others, leading to poor communication and interpersonal barriers.

7. Conclusions

After years of exploration, the mental health education curriculum in colleges and universities has received increasing attention from schools, and great progress has been made in the understanding of the curriculum and its construction. However, there are some shortcomings in the course of curriculum implementation, such as the lack of institutional, financial, and human resources support for curriculum organisation, excessive refinement of curriculum objectives, neglect of students' needs in curriculum content, and a single way of curriculum evaluation. This paper analyses the mental health of college students based on data mining algorithms, and the experiments show that under current educational conditions, it is important to adhere to a student-oriented in-depth learning orientation and emphasize the developmental and educational nature of classroom teaching.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] C. Liu, Y. Li, X. Zhao, X. Ren, and M. Tang, "Design and implementation of college students' psychological prediction system based on data mining," in *Proceedings of the 2021 4th International Conference on Information Systems and Computer Aided Education*, pp. 354–357, Dalian, China, September 2021.

[2] D. Raju and R. Schumacker, "Exploring student characteristics of retention that lead to graduation in higher education using data mining models," *Journal of College Student Retention: Research, Theory & Practice*, vol. 16, no. 4, pp. 563–591, 2015.

[3] Z. Rong and Z. Gang, "An artificial intelligence data mining technology based evaluation model of education on political

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9797213, 1 page
https://doi.org/10.1155/2023/9797213

Retraction

Retracted: Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Long, "Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5418332, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5418332, 9 pages
https://doi.org/10.1155/2022/5418332

Research Article

Analysis of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm

Juan Long

Chongqing City Vocational College, Yongchuan, Chongqing 402160, China

Correspondence should be addressed to Juan Long; 023382@cqcvc.edu.cn

Received 21 April 2022; Accepted 24 May 2022; Published 15 June 2022

Academic Editor: Chin-Ling Chen

The emergence and development of Chinese insurance companies are affected by their own unique national conditions. The modern marketing concept lags behind, lacks the practical experience of scientifically formulating marketing strategies, and insurance practitioners lack marketing knowledge and the ability to absorb modern marketing achievements to guide practice. Therefore, China's insurance industry inevitably has many problems in insurance marketing. In recent years, with the rapid development of big data (BD) technology, artificial intelligence, and machine learning in engineering and academia, relevant data models have been well developed. The advantages of the decision tree are its good robustness, full sample mining, high precision, fast implementation, fast running speed, and low implementation cost. This paper studies the application of the decision tree classification algorithm under the guidance of BD in insurance marketing planning. The running results of the decision tree classification algorithm model show what factors will affect the accuracy and recall rate of customer churn decision-making. The predicted value and scoring value of users are extracted to test the model, and the results are within a reasonable range. The running time of this model is 2,320.36 s, which is more efficient than the 34 min 25 s of traditional SAS. Therefore, the model can be put into use, and it is necessary to establish a long-term and stable relationship with customers.

1. Introduction

With the development of the insurance industry in China, marketing plays an increasingly important role in the business activities of insurance companies. How to correctly and effectively select and manage marketing channels has become one of the focuses of insurance companies [1]. The potential of the insurance market is huge, but the market competition is also fierce. Therefore, how to narrow the gap between China's insurance industry and the insurance industry of developed countries and enhance the competitive strength of its own insurance market is worthy of our deep thinking. Marketing is an important part of insurance operations. Strengthening the marketing management of the insurance market plays an important role in improving the market competitiveness of China's insurance industry [2, 3]. Because the emergence and development of Chinese insurance companies are affected by their own unique national conditions, the concept of modern marketing lags behind and lacks the practical experience of scientifically formulating marketing strategies, insurance practitioners lack marketing knowledge and lack the ability to absorb modern marketing achievements and apply them to guide practice, China's insurance industry inevitably has many problems in insurance marketing [4]. It is an urgent need for insurance companies to take integrated marketing as a breakthrough, innovate marketing mechanisms, and get rid of business difficulties. The deepening of financial system reform requires insurance companies to establish a marketing mechanism that truly adapts to the market economic system, which is the internal reason for promoting integrated marketing of insurance companies. As an important part of insurance marketing strategy, the choice of marketing channel is particularly important to China's developing insurance enterprises [5, 6]. Strengthening the marketing and marketing channel selection of Chinese insurance enterprises will not only help the improvement of their own management and technical level but also promote the formation and enhancement of people's insurance awareness and have a far-reaching impact on the sustainable development of the insurance market [7].

In recent years, with the vigorous development of BD technology, artificial intelligence, and machine learning in engineering and academic circles, the relevant data models have developed very well and fully, and the advantages of the decision tree are its good robustness, full sample mining, high accuracy, fast implementation, fast running speed, and low implementation cost [8, 9]. Therefore, this paper introduces the concept of a BD-guided decision tree classification algorithm to alleviate the problem of heterogeneous data processing in traditional data processing and meet the needs of different data source storage media, introduces the scalable BD analysis model to obtain the characteristics of users' interest migration, applies the algorithm based on the decision tree algorithm model and takes the specific user data of an insurance company as an example to build application scenarios for model training and data prediction, and innovatively introduces the value rate to classify users to solve the problems faced by the company, such as long time, low efficiency, and low accuracy in processing massive user data [10, 11].

Using the BD-guided decision tree classification algorithm to classify insurance marketing planning and find "high-quality and low-quality" customers in a large customer base has important practical significance for life insurance companies to explore the market and avoid business risks [12, 13]. Marketing in the narrow sense is the direct sale of insurance products. In the broad sense, in addition to selling insurance products, marketing also includes a series of business activities such as the development and design of insurance products, the investigation and research of the insurance market, rate formulation, after-sales service, and so on. The business philosophy of insurance companies still stays in the concept of products and promotion but does not implement the modern basic marketing concepts such as "starting from the interests of customers" and "making customers satisfied." The competition of enterprises is still dominated by the competition of insurance premium rate, which seriously distorts the principle that insurance marketing is more suitable for nonprice competition, and is also contrary to the development trend of international insurance industry [14]. The BD-guided decision tree classification algorithm is generalized, and a limited number of discrete data is used to replace continuous data. For example, the age range of the insured is from 6 to 60 years old. In order to find the risk probability of customers of different ages, we are only interested in the age stage of customers and do not need to know the specific age of customers. Eight years old represents the insured aged 6–10; 15 years old represents the insured aged 11–20; and so on [15].

This paper studies and innovates the above problems from the following aspects:

(1) An insurance marketing planning model based on a BD-guided decision tree classification algorithm is proposed. The decision tree classification algorithm is used to carry out BD theory and marketing mix strategy in marketing to achieve the purpose of marketing. There are many new types of insurance, especially life insurance, which can basically meet the needs of the insurance market. However, in terms of market demand, the promotion and innovation ideas of insurance products are narrow and single in form. Insurance companies only invest a lot of manpower, material resources, and financial resources in the design of insurance products but neglect or belittle the promotion of products, resulting in many people not knowing insurance and alienating insurance.

(2) The insurance marketing planning scheme of the BD-guided decision tree classification algorithm is contructed. The insurance marketing process guided by the BD decision tree classification algorithm is a process of seeking the balance between customers' needs and their own profits, and it is necessary to establish a long-term and stable relationship with customers. In order to achieve this effect, insurance companies need to establish a series of control measures such as analyzing the target market, implementing marketing plans and controlling the marketing process and finally achieve the profit target of insurance companies.

The paper is divided into five parts, and the organizational structure is as follows:

The first chapter introduces the research background and current situation of insurance marketing planning and puts forward and summarizes the main tasks of this paper. The second chapter introduces the related work of insurance marketing planning at home and abroad. The third chapter introduces the principle and model of the decision tree classification algorithm. The fourth chapter introduces the realization of the insurance marketing planning model of the BD-guided decision tree classification algorithm and compares the performance of the model through experiments. The fifth chapter is the full-text summary.

2. Related Work

2.1. Research Status at Home and Abroad. Honka and Chintagunta [16] proposed that with the development of China's socialist market economy and the deepening of the reform of the economic system, social security system, and education system, the insurance demand tends to be diversified, the group business is relatively reduced, and the decentralized business is greatly increased [16]. Putz et al. [17] proposed that insurance companies do not pay attention to analyzing the market environment, alienate policyholders, and only focus on developing new business channels, resulting in a decline in the public's trust in insurance agents [17]. Amoako and Okpattah [18] pointed out that the emergence and development of Chinese insurance companies are affected by their own unique national conditions, the concept of modern marketing is still relatively indifferent and lacks of practical experience in scientifically formulating marketing strategies, insurance practitioners lack marketing knowledge and lack the ability to absorb modern marketing achievements and apply them to guide practice, and China's insurance marketing is destined to have one or another problem [18]. Aksoy [19] proposed that traditional marketing is to invest capital, acquire raw materials, design and produce products in advance, establish sales channels, sell products, and provide after-sales service. It is a traditional form of the value chain. Insurance marketing is the reverse operation of the traditional value chain. Its core mechanism is to take the needs of customers as the starting point, provide corresponding products and services, and establish marketing channels with their own characteristics, so as to form their own products and advantages [19]. Xia et al. [20] pointed out that the overall quality of insurance marketing personnel is not high and the service level of the industry is low, which has seriously damaged the reputation of the insurance industry. Due to the lack of professional ethics or insurance and related knowledge of some employees, there are often violations such as misleading statements, premium rebates, and malicious solicitation when promoting insurance [20]. Karimi et al. [21] proposed that under the circumstances of changing market demand and increasingly fierce market competition, insurance enterprises not only can meet the production and operation of existing products but must constantly develop new types of insurance to meet the needs of customers and market competition, as well as the need to expand market share [21]. Karimi et al. [21] put forward that with the advancement of economic system reform and social and economic development, insurance demand is diversified. Especially, with the development of the private economy and the increase of individual industrial and commercial households, the social demand for insurance is also evolving towards diversification, and the consumer groups are also evolving towards diversification, and the proportion of group consumption is decreasing year by year. Faced with this development trend, marketing is changing in a diversified direction [22]. Wang et al. [23] put forward that the marketing of insurance is the basic goal of selling its insurance products, and more importantly, it also includes various service means derived from insurance products and ideas between products and services, with the ultimate goal of maximizing the value of customers [23]. Nancy et al. [24] put forward that the concept of modern marketing is relatively weak, lacking the time and experience to scientifically formulate marketing strategies, which leads to the lack of marketing knowledge and the ability to absorb modern marketing achievements and apply them to guide practice [24]. Piao et al. [25] put forward that the purpose of insurance marketing is to maintain long-term business relations with customers and to develop through continuous performance growth. Therefore, insurance marketing should select valuable customers through screening and, on this basis, create a long-term win-win relationship with customers through effective means [25].

2.2. Research Status of Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm. This paper studies insurance marketing planning through the BD-guided decision tree classification algorithm. The realization of the marketing objectives of insurance enterprises needs the support of an efficient marketing organization. High performance is accomplished by high-quality business personnel, which is also the result of efficient marketing organization operation. For the decision tree algorithm, there are a large number of features in large data sets, including some redundant, low-quality, and even irrelevant features. Their processing not only consumes a lot of computing resources, leads to the bloated structure of the tree but also risks the accuracy of prediction. Take the customer database of a life insurance company as the data source. A total of 9,238 records of policy information of all customers who opened accounts since 1999 and 1,484 records of claim information of additional insurance claims of the above customers as of October 2005 were extracted. In order to make insurance enterprises have a qualitative leap and a good marketing result, we must improve the existing insurance marketing organization structure and make a fast and efficient response to the market, so as to better provide customers with satisfactory service quality. According to the needs of consumers, design and develop insurance products, effectively adjust marketing strategies, mobilize the enthusiasm and replication quality of marketing personnel, make consumers satisfied with insurance products and services, and enhance consumers' recognition and trust in insurance companies, so as to occupy a dominant position in the fierce insurance market competition. There are 4,940 records in the processed form. Data reduction: the methods of dimension reduction, discretization, and concept layer are adopted to reduce the data. Delete the attributes in the original attribute set that are not related to the mining task to reduce the dimension.

3. Principle and Model of Decision Tree Classification Algorithm

The training process of the decision tree is iterative splitting and growing from the root node to the bottom, layer by layer. This process seems simple, but if a single layer of iteration is split from the bottom, it will have high space complexity and memory consumption. The flow classification algorithm in the BD environment needs to effectively improve the training speed of the model and reduce resource consumption. The decision tree classification algorithm is a typical classification algorithm, which can analyze and process training data, summarize data rules, generate a decision classification model, and then use this model to analyze and process new data. The high sample size of BD directly leads to the fact that the decision tree cannot be completely built in memory, and it also consumes a lot of operation time. In order to process these data efficiently, a reasonable way is to use the distributed storage and bandwidth resources of the cluster and use tasks parallel methods such as MapReduce or MPI to reduce the computing load. Basic data processing: the original user data is entered to form a basic data lake, and the data is imported into HBase and Oracle databases. The data is preprocessed and selected for table selection, key selection, and connector layer selection to match the data model. The basic wide table is formed through the preliminary basic processing cleaning and screening of the data table. On the basis of forming the wide table, the data washer is standardized; the sample data is simply described and counted; and the missing values are processed and standardized. The process of constructing a decision tree classification algorithm is similar to the behavior pattern of people making decisions. Overfitting will not only affect the accuracy of prediction but also make the decision rules complicated and difficult to understand. The structure of the decision tree is mainly determined by the measurement index of impurity and the method of postpruning. In the early research, the common methods adopted in these two aspects were introduced and evaluated, and the experimental results on some data sets showed their influence on decision tree size and prediction accuracy. When the initialization of the decision tree classification algorithm model is completed, the window is an empty queue. At this time, the decision tree classification algorithm will start to execute the classification module and the evaluation module. The algorithm reads the input test data stream, predicts every instance in the test data stream in real time through the established decision tree model, matches the predicted result with the premade class label, and then inserts the matched result into the ADDS window to detect the concept drift. The detailed growth process of the decision tree classification algorithm is shown in Figure 1.

The process of constructing a decision tree classification algorithm is similar to the behavior pattern of human decision-making. Given a data set *s*, which contains the values of multiple attributes and the classification to which it belongs, the first thing to do is to use some statistical methods to select an attribute *a* as the root node and divide the data set *s* into multiple subsets according to the value of attribute *a*. The tree branch and leaf node of the decision tree represent the test output and class label of data attributes, respectively. Therefore, the decision tree can be designed as the structure of the flow chart, and the classification of test data can be obtained by traversing the root and leaf nodes of the tree. The purpose of the decision tree algorithm is to find the classification rules contained in the data. Its core content is to construct a decision tree with high precision and small scale. A single decision tree classifier has limited ability to deal with problems and is prone to overfitting. The basic idea of ensemble learning is to use multiple algorithms or multiple classifiers to produce a better result. Pruning refers to replacing some subtrees with direct leaf nodes, and the class of leaves is represented by the class of most training samples contained in the replaced subtree. It can reduce the complexity of tree structure, improve the over-fitting situation to improve the prediction accuracy, quickly predict the results, and simplify the decision-making process. Back pruning was originally proposed by Breiman. Compared with the front pruning with vision effect problem, it refers to the pruning after the initial construction of a decision tree. In order to solve the problems of limited memory and training

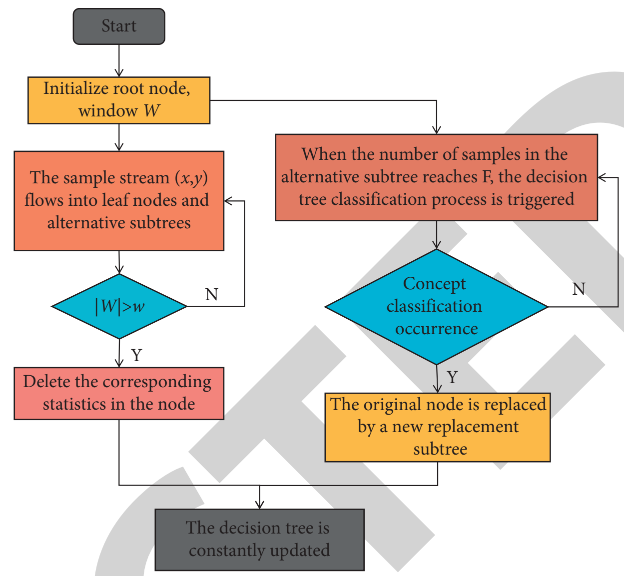


Figure 1: Flow chart of decision tree classification algorithm.

efficiency faced by the streaming data classification algorithm, the most effective way is to use the divide and conquer method to decompose the original computing task into several identical subtasks to deal with so that each computer node can balance the load. According to the idea of the divide and conquer method and the growth process of the decision tree, three parallelization strategies can be proposed: task parallelization, horizontal parallelization, and vertical parallelization. We select 1,000 samples in the training data set, and each sample contains the above 4 learning behavior attribute values, which can construct a 4-layer decision tree, and its structure is shown in Figure 2.

The classification model of the decision tree shows a series of IF-THEN rules. When a new test data enters, the data starts from the root node, goes along the branch to the leaf node, and finally gets the classification result. The data of learners' learning behavior is a series of vectors, each of which has different attribute units. IF-THEN rules can be constructed according to the model to get the classification of learning evaluation.

The memory occupation of each leaf node in the decision tree is $o(dvc)$, where c is the number of classified decisions, d represents the total number of attributes, and v is the maximum number of attribute values for each attribute. It is used to measure the ability of a given attribute to distinguish training samples. The calculation formula of information gain is shown in the following formula:

$$\text{Entropy}(S) = \sum_{i=1}^{c} - p_i \log(p_i), \tag{1}$$

$$\text{Gain}(S, A) = \text{Entropy}(s) - \sum_{v \in \text{Values}(A)} \frac{|S_V|}{S} \text{Entropy}(S_v). \tag{2}$$

Equation (1) is the calculation formula of information entropy, where p_i is the proportion of samples of different

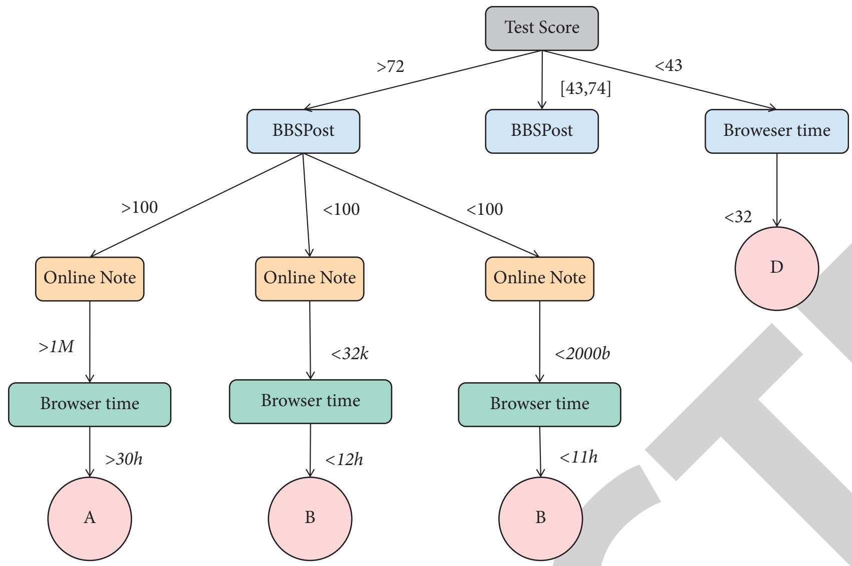


Figure 2: Classification model structure of decision tree.

categories C_i in sample set S. Equation (2) is the calculation formula of information gain, where Values(A) is the set of all possible values of attribute A and S_V is the subset of attribute A in S (i.e., $S_v = \{s \in S | A(s) = v\}$).

With the continuous entry of data flow, the fitting degree of the decision tree model is getting higher and higher; at this time, the classification accuracy of the decision tree should be more and more stable, and the difference between the classification accuracy P_{real} in the sliding window and the optimal classification accuracy P_{optimal} is also getting smaller and smaller. We define $\Delta P = P_{\text{optimal}} - P_{\text{real}}$ to detect the error between the optimal classification accuracy and the real-time classification accuracy in the sliding window, assuming that we give the confidence δ and the confidence interval h for the accuracy error, which can be obtained by the inequality in Hoeffding.

$$P(\overline{X} - E[\overline{X}] \geq h) \leq e^{-2nh^2}. \quad (3)$$

Let $\overline{X}$ be the estimated accuracy and $E[\overline{X}]$ be the true accuracy, then the meaning of the above inequality is that the probability that the deviation between the estimated accuracy and the true accuracy exceeds h is not greater than e^{-2nh^2}.

To find the most accurate Hoeffding upper limit, that is, find the minimum value on the right side of the equation have

$$s = \frac{4t}{\sum_{i=1}^{n} (b_i - a_i)^2}. \quad (4)$$

Bring s into the equation and get

$$P(S_n - E[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2}\right). \quad (5)$$

In the decision tree classification algorithm, Hoeffding inequality can be used to determine the minimum number of samples needed for the node splitting in the decision tree in a given confidence interval δ.

According to symmetry, there is

$$P(\overline{-X} + E[\overline{X}] \geq h) \leq e^{-2nh^2}. \quad (6)$$

It can be obtained by formulas (5) and (6):

$$P(|\overline{X} - E\overline{X}|) \geq h \geq 2e^{-2nh^2}. \quad (7)$$

That is, for a given confidence δ, within the confidence interval where the expected accuracy $E[\overline{X}]$ width is $2h$, there is

$$\delta \leq 2e^{-2nh^2}. \quad (8)$$

A parallel window scheme based on a decision tree classification algorithm initializes multiple windows, divides real-time sample streams into S_1 and S_2, monitors the hidden information distribution of the two streams, and obtains the formula according to the Poisson process and Hoeffding inequality.

$$P_r\{X\} \geq (1+\varepsilon)E[X] \leq \exp\{-((1+\varepsilon)1n(1+\varepsilon) - \varepsilon)E[X]\}, \quad (9)$$

where $E(X)$ is the period value of the sample flow and ε is the limit of concept drift that needs to be detected. According to the Taylor series formula and Poisson process, the following formulas can be expected:

$$1n(1+x) = \sum (-1)^{n+1} \bullet \frac{X^n}{n} = X - \frac{X^2}{2} + \frac{X^3}{3} \cdots, \quad (10)$$

$$E[Y_n] = nE[X_n] = n\lambda_x. \quad (11)$$

The threshold for judging the conceptual drift of the sample flow is $\varepsilon = \sqrt{2\lambda 1n(2/\delta)(1/n_1 + 1/n_2)}$. In this paper, the sample flow is divided into two streams S_1 and S_2; then $n1$ is the number of samples in the sample flow S_1; and n_2 is the number of samples in the sample flow S_2. δ is the confidence of Hoeffding Bounds, and λ is the mean value of samples in the sample flow.

4. Realization of Insurance Marketing Planning

4.1. Insurance Marketing Planning Based on BD-Guided Decision Tree Classification Algorithm. From the perspective of the development history of the world insurance industry, China's insurance industry has been developing for a short time. However, with the deepening of the reform and opening up and the continuous improvement and development of the market economy system, China's insurance industry has been developing in the form of a blowout, especially since the reform and opening up, and its development speed and scale have far exceeded people's imagination, and the insurance market has made remarkable achievements. The insurance marketing of the decision tree classification algorithm obtains its own BD benefits on the premise of meeting customers' needs. It requires every member of the organization to think of customers and do their best to create more value for customers. The characteristics of insurance marketing and the particularity of insurance products determine the characteristics of insurance marketing. We can summarize the marketing characteristics of insurance products as follows:

① Change potential demand into actual demand

Most people's demand for insurance is potential, and insurance products are invisible and intangible abstract goods. Most people seem to have no urgency for it, especially for life insurance products. Therefore, the insurance marketer must change the potential demand of the insured into the actual demand through active marketing.

② Turn negative demand into a positive demand

Because most insurance products are related to people's life and death, for many people, their demand for insurance products is a negative demand. That is to say, people take negative evasive attitudes and behaviors towards insurance products because they do not like or understand them.

③ Change one-way communication into two-way communication

As a marketer of insurance products, one-way communication must be changed into two-way communication. That is to say, through active marketing, the information to be conveyed by enterprises will be transmitted to consumers through information media in a way that consumers can understand and accept, and consumers' feedback on information will be tracked and paid attention to, so as to collect consumers' opinions and reactions on the provided insurance products and timely adjust and improve service strategies to achieve customer satisfaction.

Now that China has joined the WTO, in order to gain a firm foothold in the further open Chinese insurance market and remain invincible in the competition, many insurance companies use BD to meet the needs of economic development and enhance their ability to connect with the international insurance market. Modern marketing theory believes that using the decision tree classification algorithm to classify products is divided into three forms: tangible products, intangible labor services, and social behavior. In order to achieve this goal, fundamentally speaking, we should use a decision tree classification algorithm to carry out BD theory and marketing combination strategy in marketing in order to improve consumers' awareness. There are many new types of insurance, especially life insurance, which can basically meet the needs of the insurance market. However, from the perspective of market demand, the promotion and innovation ideas of insurance products are narrow, and the form is single. Insurance companies only invest a lot of human, material, and financial resources in the design of insurance products but ignore or despise the promotion of products, resulting in many people not knowing and alienating insurance. Insurance is an economic behavior in which the decision tree classification algorithm guided by BD predicts the possible uncertain events and collects the insurance premium, establishes the insurance fund, and transfers the risk from the insured to the insurer in the form of contract, and the majority participating in the insurance jointly share the loss of a few. Therefore, insurance marketing has different characteristics from other commodities. The process of insurance marketing guided by the BD decision tree classification algorithm is a process of seeking the balance between customer needs and their own profits. It is necessary to establish a long-term and stable relationship with customers. In order to achieve this effect, insurance companies need to establish a series of control measures such as analyzing the target market, implementing marketing plans, and controlling the marketing process and finally achieve the profit goal of the insurance company.

4.2. Experimental Results and Analysis. In this experiment, the data selection basically processes the row and column dimension data of the wide table. Because a wide table with about 25 attribute columns is generated from the data in the actual process, the data selection can avoid the disaster of high-dimensional data in data processing, and some data are normalized in the data processing process to adapt to the matching degree of the model as shown in Table 1.

It can be concluded from Table 1 that the running platform of this model is based on the Hadoop distributed file system, and its good high fault tolerance and high throughput data access are more suitable for the application of large-scale data sets. The application environment of this model is based on the basic running environment of HDFS, using Python data processing language, and the operating system version is centosrelease6 5 (final), set up 6 clusters. Relevant information of each device are as follows: Intel (R), e5606, @ 2.13 GHz, 2128.000 mHz, and cache size: 8,192 kb.

The running results of the decision tree classification algorithm model show which factors affect the decision of customer churn, and more valuable customer information

Table 1: Description and statistics of sample data.

Customer value rate category	Number of samples (PCs.)	Average age (years)	First year annualized premium	Income
4	1,251	43.0045	7,335.0658	104,969.65
3	5,381	38.1036	8,051.2526	170,750.08
2	924	41.4388	3,437.5842	81,722.66
1	277	45.9191	3,310.5149	99,677.58
0	10,456	49.1984	2,846.6048	46,312.29

Table 2: Full sample data operation model.

TN	FN	FP	TP	TestErr	Recall	Precision
363,235	27,784	11,216	57,168	0.0849024	0.67291584	0.67291569

can be obtained through evaluation. The evaluation methods include accuracy rate, recall rate, PR, ROC, and so on. Among them, the real TP: the sample type is correctly classified by the data model to predict the number of hits in the right class; FN: the sample type is misjudged by the data model as the quantity of other types; FP: the number of samples that do not belong to the correct category and are misjudged as the correct category by the data model; and true negative (TN): the sample type belongs to the correct category and is misjudged by the data model as the quantity of other types. The results of the whole sample data running model are shown in Table 2.

According to Table 2, the accuracy and recall rate are used here to extract the predicted value and scoring value of users for the model test. The obtained values are within the reasonable value range. The running time of the model is 2,320.36 s, which is more efficient than the 34 min 25 s run out of the traditional SAS. Therefore, this model can be put into use. In this experiment, in the field of agricultural economics, many foreign scholars have done a lot of empirical research on farmers' attitudes towards risk, especially in developing countries, and generally put forward the conclusion that farmers are risk-averse. In order to make this empirical conclusion in line with the reality of contemporary Chinese farmers, we visited the farmers in Fengshu village, Low Ping village, Liangshan village, and Baoyuan village in the Hongtang area of Xiangxiang City. Three surveys were conducted and compared in this experiment. The experimental results are shown in Figures 3–5.

As can be seen from Figures 3–5, the average importance of risk preference is about 8.5%; the average importance of risk aversion is about 6.3%; the average importance of risk aversion is about 4.6%; the average importance of risk aversion is about 10.2%; the average importance of risk aversion is about 7.1%; and the average importance of risk aversion is about 4.6%. Generally speaking, older people tend to avoid risks, and the proportion of young farmers who tend to take risks is much higher than that of older people.

In this experiment, the matching between the premium income of China's agricultural insurance and that of property insurance from 2015 to 2021 was analyzed, and two experimental investigations were conducted to compare them. The experimental results are shown in Figures 6 and 7.

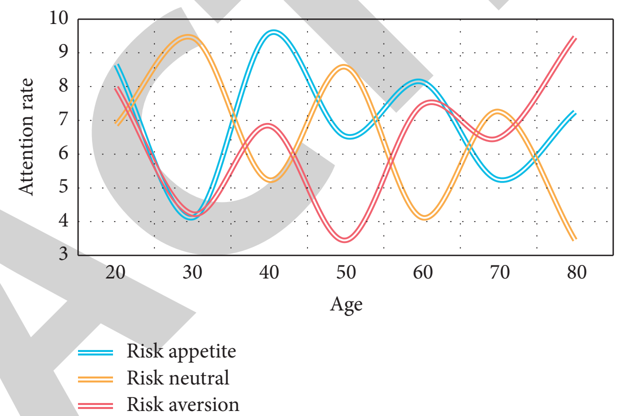


Figure 3: The farmers' attention to risks from the age structure.

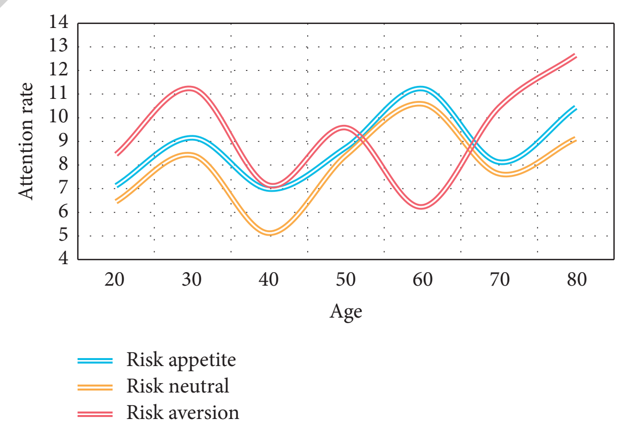


Figure 4: The farmers' attention to risks from the age structure.

As can be seen from Figures 6 and 7, the premium of China's agricultural insurance premium income in property insurance premium income from 2015 to 2017 is generally too low. Before the implementation of the policy pilot, the premium did not exceed 8,500. After the implementation of the policy pilot, although there was an obvious increase, it was less than 9,500. China's agricultural insurance premium income from 2018 to 2021 does not match the development of property insurance premium income. Before 2019, the property insurance premium increased by a large margin year by year, while the agricultural insurance premium has

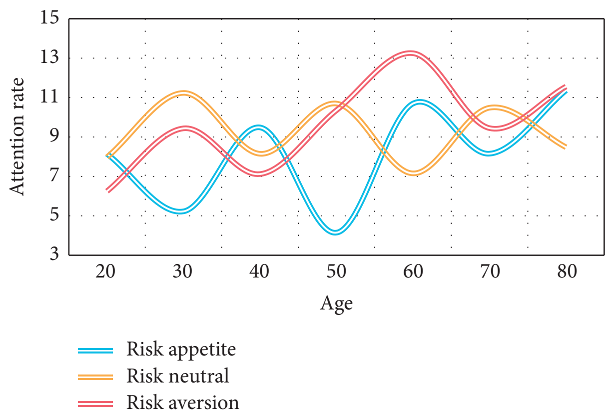


Figure 5: The farmers' attention to risks from the age structure.

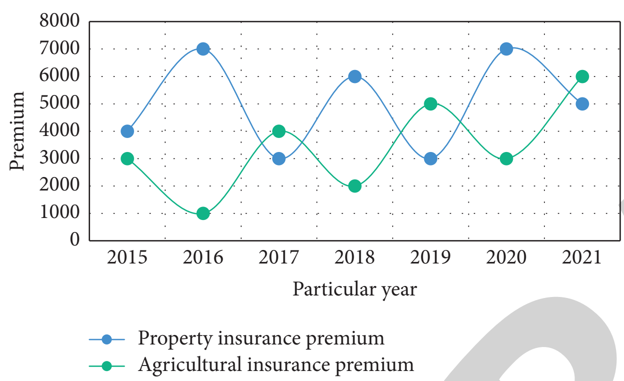


Figure 6: Trend correlation chart of agricultural insurance premium income and property insurance premium income in China.

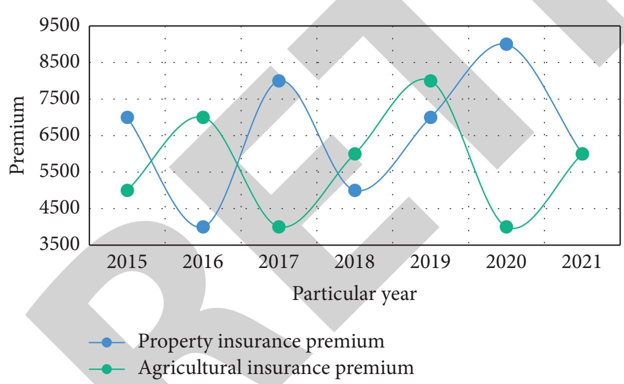


Figure 7: Trend correlation chart of agricultural insurance premium income and property insurance premium income in China.

been at a very low level year later, the development trends of the two are relatively close, but in terms of quantity, the premium income of agricultural insurance is still too low, and the premium income of property insurance is too low.

5. Conclusions

Through the research on the basic characteristics of decision tree classification algorithm, the combination point of decision tree classification algorithm and insurance user churn rate is found, and the big data model of insurance marketing planning is established; The model is based on open source HDFS environment and has good scalability. Insurance agents play an important role in obtaining decentralized insurance consumption business, supplementing the sales capacity of insurance companies, and promoting the construction of a marketing network of insurance companies and have become an important channel for the business source of China's insurance market. According to the nondirectional classification of the nature of employment, insurance agents can be divided into personal agents, part-time agents, and professional agents. After selecting the target big data market, insurance companies under the decision tree classification algorithm should design different insurance types and marketing schemes for each target market to meet the insurance needs of different consumers. In this paper, for the big-data-guided decision tree classification algorithm, the use of conventional processing methods in data processing will inevitably lead to the loss of some data and the decline of prediction accuracy. The insurance marketing planning of the decision tree classification algorithm guided by big data is studied. The operation results of the decision tree classification algorithm model show which factors affect the decision-making of customer churn. The accuracy and recall rate are used to extract the predicted value and scoring value of users for the model test. The values are within the range of reasonable values, and the operation time of the model is 2,320.36 s, Compared with the 34 min 25 s of traditional SAS, it is more efficient. Therefore, this model can be put into use. Insurance is an economic behavior that predicts possible uncertain events and collects insurance premiums through big-data-guided decision tree classification algorithm, establishes an insurance fund, and transfers risks from the insured to the insurer in the form of contract, and the majority participating in insurance jointly share the losses of a few. Therefore, insurance marketing has different characteristics from other commodities.

Data Availability

The labeled datasets used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the 2020 Chongqing Municipal Education Commission Humanities and Social Sciences (No. 20SKGH337), 2020 Chongqing City Vocational College (No. XJSK202001009), and Chongqing City Vocational College Smart Retail Collaborative Innovation Centre.

References

[1] J. Hong and S. H. Seog, "Life insurance settlement and the monopolistic insurance market," *Insurance: Mathematics and Economics*, vol. 81, pp. 36–50, 2018.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9865853, 1 page
https://doi.org/10.1155/2023/9865853

Retraction

Retracted: Design of College Scheduling Algorithm Based on Improved Genetic Ant Colony Hybrid Optimization

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Li, Q. Xie, and H. Zhang, "Design of College Scheduling Algorithm Based on Improved Genetic Ant Colony Hybrid Optimization," *Security and Communication Networks*, vol. 2022, Article ID 2565639, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2565639, 13 pages
https://doi.org/10.1155/2022/2565639

Research Article

Design of College Scheduling Algorithm Based on Improved Genetic Ant Colony Hybrid Optimization

Ting Li,[1] Qiang Xie,[2] and Hua Zhang[3]

[1]Center for Evaluation and Faculty Development, Hunan First Normal University, Changsha, Hunan 410205, China
[2]Academic Affairs Office, Hunan First Normal University, Changsha, Hunan 410205, China
[3]School of Computer Science and Engineering, Hunan University of Information Technology, Changsha, Hunan 410151, China

Correspondence should be addressed to Hua Zhang; zhanghua@hnuit.edu.cn

Received 18 April 2022; Revised 16 May 2022; Accepted 30 May 2022; Published 14 June 2022

Academic Editor: Chin-Ling Chen

With the gradual expansion of college scale, the professional categories in colleges and universities are becoming more and more complete, and the volume of courses is becoming more and more huge. In the meantime, the number of students is growing by leaps and bounds, and the teaching resources are subject to more and more complicated teaching tasks. The workload and the difficulty of scheduling in teaching management are also on the rise year by year. This paper proposes a design of a college scheduling algorithm based on an improved genetic ant colony hybrid optimization algorithm. Firstly, the fitness-enhanced elimination law is proposed to improve the selection process of traditional genetic algorithms. Subsequently, the gene infection crossover method is proposed to ensure the increase of the average fitness value in the evolutionary process. Next, the unnecessary replication operation in the traditional genetic algorithm is removed to enhance the operation speed of the algorithm. Finally, the parallel mechanism of fuzzy adaptive is introduced to improve the convergence and stability of the algorithm. For the ant colony optimization algorithm, a nonuniform pheromone distribution is used according to the position of the current raster relative to the starting point, which makes the initial pheromone concentration of the dominant raster higher and avoids blind search by ants. The ant movement rules are redefined by the directional neighborhood expansion strategy to further shorten the path. The experimental results indicate that the hybrid optimization algorithm outperforms other algorithms in terms of performance in terms of scheduling success and scheduling time, and it can be applied in practical scheduling because of the high quality of courses schedule.

1. Introduction

Education is a fundamental part of the sustainable development of the country and society, and it is also a long-term basic national policy of China to develop the country through science and education. Education is an important means to improve the education level of society and the quality of the nation, especially higher education. Hence, China's colleges have been expanding their enrollment scale each year and continuously increasing the number of people receiving higher education [1, 2]. In the meantime, the level of scientific research and management is constantly improved and gradually transformed from quantitative change to qualitative change. The process of education and training of talents with high precision technology and skills is long and strict, and each link needs reasonable planning and arrangement. With the continuous refinement of college majors, colleges and universities are constantly developing and innovating in the talent training system. In terms of course categories and link settings, they all strive to cultivate personalized talents with both professional depth and professional span, which undoubtedly has new requirements and challenges for the scheduling work in colleges and universities [3–5]. In-depth research and discussion on the scheduling system of colleges and universities is the basic work of implementing the national talent cultivation strategy and plays a significant role in the teaching management of colleges and universities.

Scheduling work is essentially the optimal use of various teaching resources in universities to ensure that basic teaching can be carried out in an orderly manner [6]. It is a proposition that needs to solve how to arrive at an optimal solution under multidimensional constraints. That is, various constraints such as time, space, and people need to be satisfied simultaneously in the system design to ensure that no conflicts arise between various constraints in the process of classroom teaching [7]. Avoiding various types of conflict problems such as the same batch of students needing to appear at different teaching sites at the same time or a particular classroom teacher being assigned different teaching tasks at the same time results in the classroom teaching work not being carried out properly [8, 9]. The more humane demands such as not too many hours of classes in a day for one teacher and not too many hours of classes in the same course for the same batch of students are also a challenge to the newer scheduling work. At present, the traditional manual scheduling method commonly used in domestic universities is mainly based on the previous scheduling experience. However, because the traditional manual scheduling method does not have complete systematic theoretical support, let alone data modeling by computer, it often has disadvantages in actual operation such as multiple involvement, large communication volume, slow speed, low efficiency, and error-prone and strong subjective consciousness. This approach is obviously no longer applicable to contemporary higher education institutions. For multiple departments and multiple course categories (basic courses, specialized courses, compulsory and elective courses, etc.), a large number of students need to complete various types of teaching schedule arrangements each semester. If the course arrangement is carried out manually, the workload is so large that it will definitely take a lot of effort and time of teaching administrators, and it is difficult to give a better schedule that is reasonable and feasible and can make full use of teaching resources [10, 11].

Multidimensional information collection and high-speed data processing are realized and gradually popularized and applied currently. Before the scheduling process in colleges and universities, the courses are mathematically modeled, and constraints such as courses, instructors, teaching locations, and number of students are set, and computer algorithms are used to complete the solution in order to realize computer-aided decision-making automated scheduling. It not only solves the rationality and feasibility of course arrangement in space and time and improves the efficiency of scheduling work, but also enhances the level of informationalization of teaching management, which is of great importance.

Algorithms for computer-aided decision scheduling systems are commonly used in domestic universities presently. These methods include graph coloring [12], which transforms the university course scheduling problem into a graph with vertices representing courses and edges representing constraints. The number of colors corresponds to feasible time slots. The graph coloring method assigns a finite number of colors to vertices, and no two adjacent vertices that are linked by an edge are of the same color. Genetic Algorithms (GA) and improved GA [13–15] assess course satisfaction with penalty function by genetically coding course scheduling constraints. Linear programming [16], simulated annealing (SA) algorithm [17], taboo search (TS) algorithm [18], etc., these algorithms are very popular now. The shortcoming of these algorithms is that they have difficulty in dealing with the constraints of the course scheduling process and can only produce feasible solutions with unsatisfactory results. There are also course scheduling methods on deep learning [19, 20], but the computational complexity of deep learning is high and takes longer time.

Several studies have shown that hybrid algorithms show better results in solving college course scheduling problems. Integration of local search algorithm (LS) into particle swarm optimization algorithm (PSO) [21, 22] was used to construct the optimal solution for college course scheduling. Constraint propagation is integrated with genetic algorithm to obtain the approximate optimal solution for college course scheduling. The disadvantage of genetic algorithm is the long computational time. The particle swarm optimization algorithm converges faster than the genetic algorithm and does not require much parameter adjustment. PSO alone cannot solve the constraint satisfaction problem. The curriculum scheduling problem is a constraint satisfaction problem, so it is necessary to find a way to deal with the constraint conflicts in the control curriculum scheduling problem.

This paper proposes an improved genetic ant colony based hybrid optimization algorithm for college scheduling. The algorithm gives constraints on the scheduling problem and applies the improved genetic ant colony hybrid optimization algorithm to solve the scheduling problem. The empirical verification demonstrates that the proposed algorithm achieves good results in practical scheduling problems.

Section 2 of this paper is the state of the art of the problem of scheduling. Section 3 is the methodology of proposed algorithm. Section 4 is the implementation of the scheduling algorithm. Section 5 is the result analysis and discussion, and Section 6 is the conclusion.

2. State of the Art

The core task is to arrange courses, classes, teachers, and classrooms without conflicts in each lecture period and to ensure that they meet the constraints set by teachers in advance. The introduction of the class system in universities has made the scheduling problem more complicated. The constraints of class scheduling are further increased, and the lack of teaching resources in schools is further highlighted.

The scheduling problem is described as follows: the courses set $SC = \{s_1, s_2, \cdots, s_s\}$. The class set $CC = \{c_1, c_2, \cdots, c_c\}$. The set of teachers $TC = \{n_1, n_2, \cdots, n_n\}$. The classroom set $RC = \{r_1, r_2, \cdots, r_r\}$. The time period set $PC = \{u_1, u_2, \cdots, u_u\}$. Before scheduling a course, it is necessary to set up a schedule for different grades and classes and to determine the relationship between courses, classes, teachers, and classrooms.

There are two types of constraints to consider when scheduling classes. The hard constraints are the conditions that must be followed in the scheduling process, and the schedule can only be arranged in accordance with the hard constraints to ensure that the scheduling resources do not conflict with each other, as follows:

(1) Only 1 course can be scheduled for the same class in the same teaching period:

$$\sum_{w=1}^{CC}\sum_{x=1}^{PC}\sum_{y=1}^{SC} c_w u_x s_y n_t r_z \leq 1. \tag{1}$$

Formula (1) indicates that in the same class c_m in the same class period p_i only 1 course can be scheduled at most s_y, by the teacher n_t in the classroom r_z classroom.

(2) Only a maximum of one course can be scheduled in the same classroom during the same teaching period:

$$\sum_{z=1}^{RC}\sum_{x=1}^{PC}\sum_{y=1}^{SC} c_w u_x s_y n_t r_z \leq 1. \tag{2}$$

Formula (2) denotes the same classroom r_z in the same lecture period u_x only 1 course can be scheduled at most s_y, by the teacher n_t in the class c_w class.

(3) A maximum of 1 course can be scheduled by the same instructor in the meantime, i.e.,

$$\sum_{t=1}^{TC}\sum_{x=1}^{PC}\sum_{y=1}^{SC} c_w u_x s_y n_t r_z \leq 1. \tag{3}$$

Formula (3) indicates that the same instructor n_t in the same class period u_x can only be assigned at most 1 course s_y and in the classroom r_z for the class c_w classroom.

Soft constraints are nonmandatory rules before scheduling a class. These rules are not necessary to be met, but they can have a significant impact on the rationality of the schedule and user satisfaction, as follows:

① The weekly class schedule of the same course is spread out as much as possible

② If a course is scheduled to have a priority, the course should be scheduled in the session that has the highest priority, such as the main course in the morning

③ Teacher continuity setting is the maximum number of consecutive lessons that can be taught by one teacher

④ Some courses are offered and others are not scheduled afterwards

3. Methodology

3.1. Enhanced Infection Genetic Algorithm

3.1.1. Enhancement of the Law of Elimination. Conventional genetic algorithms have small differences in fitness values among individuals in the late evolutionary stage, and the selection process is weakly competitive, resulting in stagnant population evolution, low accuracy in finding superiority, and slow convergence. After obtaining the individual fitness value, the average fitness of the population is calculated and then the average fitness of the individual and the population is compared. This operation enables quickly eliminating more individuals with low fitness in the early stage of population evolution. The culling criterion varies with the number of generations. At later stages of evolution, this enhances competition among similar individuals and prevents the population from stagnating. The mathematical description of the fitness value enhanced elimination law is as follows.

Let there be n individuals in a population, and the phenotype of the xth individual in the ath generation is x and the fitness value is $f_a(i_x)$. The mean fitness value of the ath generation is $1/t\sum_{x=1}^{t} f_a(i_x)$. The recalculated fitness value of the xth individual is

$$f_t(i_x) = \begin{cases} \left(f_a(i_x) - \dfrac{1}{t}\displaystyle\sum_{x=1}^{t} f_a(i_x)\right)^z, & f_a(i_x) > \dfrac{1}{t}\displaystyle\sum_{x=1}^{t} f_a(i_x), \\ 0, & f_a(i_x) \leq \dfrac{1}{t}\displaystyle\sum_{x=1}^{t} f_a(i_x), \end{cases} \tag{4}$$

where z is the reinforcement competition coefficient. z directly affects the evolution speed and the accuracy of the algorithm for finding the best performance; in order to study the influence of z value on the value of individual fitness, the curves are plotted based on $j = i^z$ $(i \geq 0)$ function based on $z = 0.3$, 0.2, 1, 2, 3, respectively.

From the experimental results it is clear that the function y is convex when $0 < z \leq 1$. And as the value of k decreases, the curve trend tends to flatten. When $z > 1$, the function y is concave and the curve tends to steepen as the value of z increases. It can be seen that, for the enhanced elimination law, a smaller value of z decreases the difference in fitness values between individuals and retards population evolution. A larger value of z increases the difference in fitness values between individuals and accelerates population evolution. Therefore, the larger the z value, the greater the difference in fitness values among similar individuals at the later stage of population evolution, the more intense the competition, and the better the search accuracy.

3.1.2. Genetic Infection Crossover Method. Conventional genetic algorithms use crossover and mutation for random search. When the chromosomes of two individuals are crossed over to produce new individuals, this approach is likely to produce offspring with lower fitness values when the fitness value of the parent individual is larger, thus reducing the average fitness of the population and deviating from the optimal solution. In contrast, gene infection crossover effectively prevents the reduction of fitness of new individuals by replacing most of the genes of the

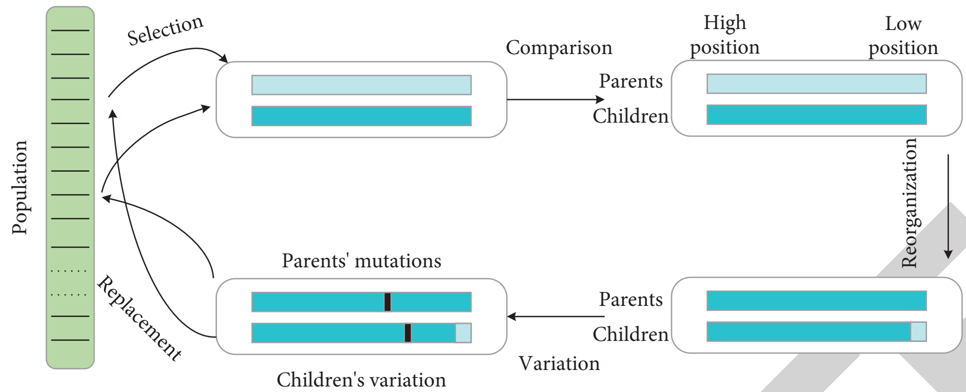


Figure 1: Genetic infection cross-operation.

parent with those of individuals with low fitness. As shown in Figure 1, the gene infection crossover method is described as follows:

Step 1. The fitness values of individuals are recalculated using the augmented elimination law such that individuals with lower fitness are not selected when the selection operation is performed.

Step 2. For each individual in the selected population, a parent is selected according to the principle of proportional selection, and the fitness of the individual is compared with the fitness of the parent. When the individual's fitness is less than the parent's fitness, a lower point in the parent's chromosome is selected and the corresponding gene segment in the individual's chromosome is replaced with the higher gene infection before this point. When the individual's fitness is greater than the parent's, no further crossover of gene infection is performed.

In the linear problem, when the parent replaces most of the high genes with infected individual genes, the individual's phenotype will rapidly approach the parent, preventing the generation of individuals with lower fitness values and thus preventing the reduction of the average fitness value of the population. Moreover, by retaining their own low genes, individuals retain a certain search ability and are able to search near the parent's phenotype, increasing the probability of searching for the optimal solution. In contrast, in nonlinear problems, a search based solely on the parent's phenotype can easily lead to a local optimal solution. Therefore, when performing gene crossover, the principle of proportional selection is applied to each individual to select the parent to ensure that the parental phenotypes can be adequately selected and compared. This ensures the speed of search, while preserving the diversity of parental expressions and preventing the emergence of local optimal solutions.

3.1.3. Enhanced Infection Genetic Algorithm

Step 1: the algorithm first sets parameters such as crossover rate and variance rate and initializes the population. Initialize the population; i.e., based on the selected coding method, a coding set with completed individual coding is generated based on the selected population size as the initial value for the algorithm to run. The commonly used encoding methods are binary encoding and gray code encoding.

Step 2: calculate the new fitness value. That is, the fitness value is calculated based on the original fitness function, and then the fitness value of individuals in the population is updated by the enhanced infection rule.

Step 3: determine whether the conditions for stopping evolution are satisfied. When the conditions for stopping evolution are satisfied, exit the program and give the result; otherwise proceed to the next step. Commonly used control conditions of the algorithm are evolutionary algebra, planning error, etc.

Step 4: perform gene infection crossover operation, and after that, perform mutation operation to move to step 2. Using the principle of proportional selection to crossover individuals of the population, the evolutionary speed is enhanced while ensuring the computational accuracy.

Compared with the conventional genetic algorithm, the enhanced infection genetic algorithm is improved in two main aspects: first, after obtaining the fitness value of each individual, the fitness value is recalculated according to the enhanced elimination law. It enhances the competition of different genes, increases the elimination of individuals with small fitness values in the early evolutionary stage, and speeds up the convergence rate. It increases the competition of similar chromosomes in the late evolutionary stage and improves the convergence accuracy. Second, by improving the crossover method of gene infection, it ensures that the fitness value of offspring individuals is higher than that of their parents, avoiding the generation of individuals with lower fitness values and preventing the decrease of the average fitness value of the population.

3.2. Adaptive Parallelism Mechanism. The introduction of parallel mechanism in genetic algorithm can effectively combine the natural parallelism of GA and the fast concurrency of computer, which is an important direction to improve the performance of algorithm proposed in recent

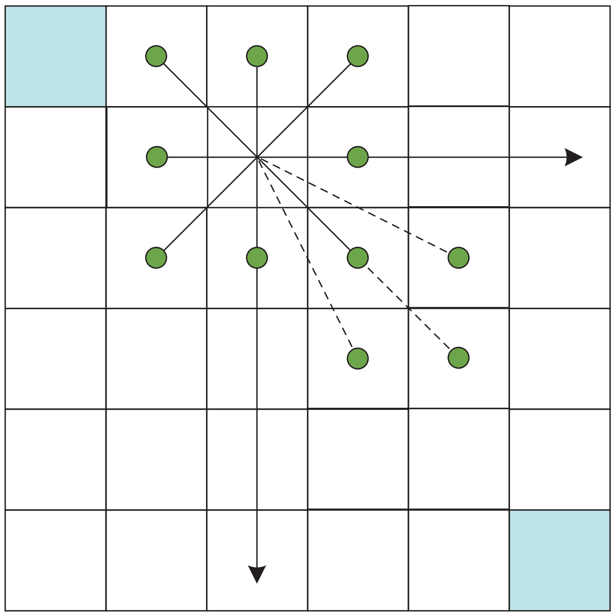


Figure 2: Feasible raster after directional neighborhood expansion.

years. Currently, there are four basic types of parallel mechanism models for genetic algorithms: Master-slave model, fine-grained model, coarse-grained model, and hybrid model. The parallel mechanism proposed in this paper adopts a coarse-grained model.

3.2.1. Evolutionary Strategies for Parallel Mechanisms. The existing genetic algorithm improvements can be summarized as coding, microgenetic strategy (improvement of genetic operator and parameter selection), and macrogenetic strategy (improvement of algorithm operation mechanism). In this paper, macroscopic multigroup parallelism mechanisms are studied rather than specific evolutionary strategies. Specifically, the larger value of P_m and the smaller value of P_c are called the exploration strategy, and the genetic operation is chosen to perform the mutation operation first and then the crossover operation, and finally the selection replication operation is performed. This evolutionary strategy facilitates the population to jump out of the local optimum and explore the new solution space. The smaller value of P_m and the larger value of P_c are called development strategy. This strategy first performs crossover operations, then mutation operations, and finally selective replication operations. This strategy can recombine existing genes to produce new individuals in anticipation of better trait changes. The normal strategy uses P_c and P_m, which are in between the developmental strategy and the exploration strategy, and uses the GA selection replication, crossover, and mutation operations in the evolutionary operation, and this strategy has the functions of both. In addition, the parallel mechanism of the evolutionary strategy of GA uses the public population as a container to continuously receive outstanding individuals passed from other populations and further develop them. First, the definition of evolutionary potential is given:

Definition 1. The evolutionary potential of individual x_i exhibited by local search is defined by formula (5):

$$V(i_x) = \omega\left(1 - Z^{\lambda}\right)\left(f'(i_x) - f_{\max}\right), \tag{5}$$

where ω denotes the weight of evolutionary potential. Z denotes the temperature decay coefficient. λ denotes the number of new solution searches. $f'(i_x)$ denotes the optimal solution of chromosome x_i after local search. $f_{\max}$ denotes the population fitness maximum. The evolutionary potential of an individual is related to the number of searches for new solutions and the fitness of the optimal solution. The lower the number of searches, the higher the evolutionary potential.

The public population first performs crossover operations on the best individuals. After the crossover is finished all individuals are searched and evolutionary potential is calculated according to the local search strategy, and then the fitness is updated according to formula (6). Finally, the chromosomes with the population size number are selected according to the fitness into the next generation.

$$f(i_x) = f(i_x) + V(i_x). \tag{6}$$

The purpose of a public population is to enable the population to explore the solution space around individuals, to explore individuals with evolutionary potential, and to discover new solutions. Individuals with high evolutionary potential have a higher probability of being selected for the next generation. On the one hand, it can enrich the population diversity and maintain the search range of the population, and on the other hand, crossover with other good individuals can hopefully produce new and better individuals.

3.2.2. Probability of Adaptive Determination Strategy Transformation. If the population evolution process is in stagnation or caught in the local optimal solution, the evolutionary strategy should be changed according to the population evolution. The evolutionary algebra A_f, in which the optimal solution is maintained constant, is introduced to measure the evolutionary state of the population, and the probability of the population changing its strategy is calculated by the following formula:

$$U_{cb} = \frac{A - A_n}{A} - \frac{1}{1 + exp\left[\beta\left(2A_f/A_{\max} - 1\right)\right]}, \tag{7}$$

where A_n is the current evolutionary generation. A is the total evolutionary generation. β is the control parameter with a value of 6 in the text. A_f indicates the number of generations of population stagnation. $A_{\max}$ is the maximum number of generations of stagnation.

The nonlinear probability can more reasonably control the population to change the evolutionary strategy. U_{cb} grows slowly in the early stage of evolutionary stagnation, allowing some time for the population to escape from stagnation by relying on the existing evolutionary strategy. As the number of generations of stagnation increases, the algorithm determines that the population is in stagnation.

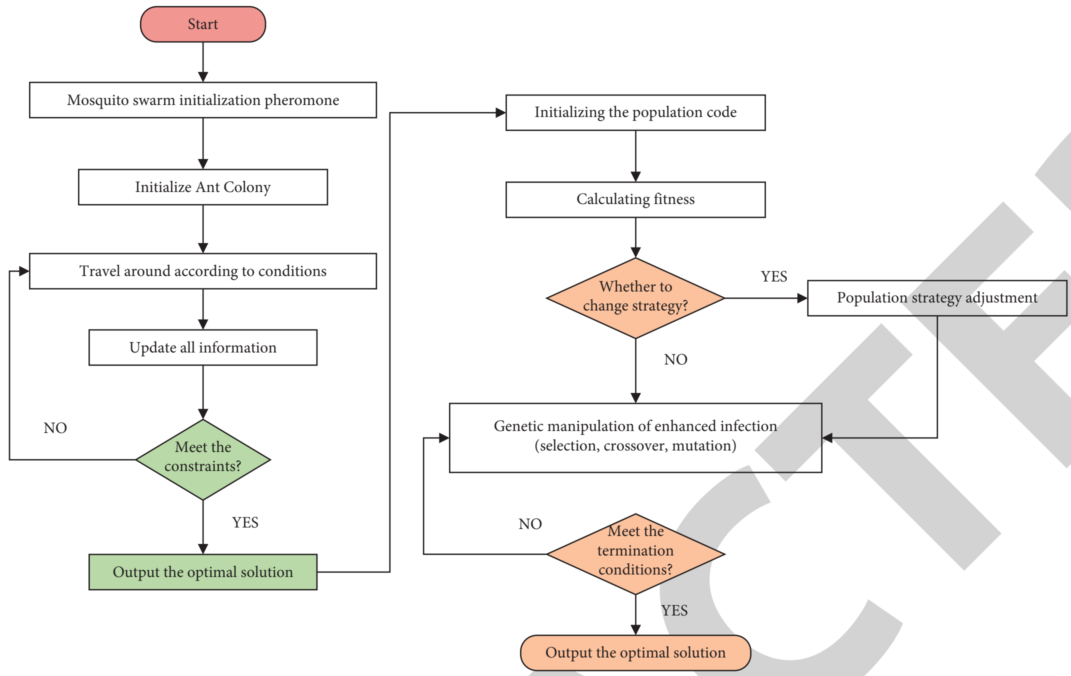


Figure 3: Improved genetic ant colony algorithm flow.

Compared with the probability of linear control, U_{cb} is significantly improved to avoid the population wasting time on the existing evolutionary strategy. Finally, to maintain a balance between population exploration and algorithm convergence rate, U_{cb} shows a linear decrease with further increase in evolutionary generations when the number of stagnation generations is constant.

3.2.3. Evolutionary Strategy of Fuzzy Inference. This paper draws on the idea of fuzzy control and applies fuzzy inference to adjust the evolutionary strategy of the population according to the individual variability and evolutionary status of the population. Formula (8) is used to represent the evolutionary status of the population.

$$E_1 = \frac{f_{\max} - f_{avg}}{f_{\max}} \in [0, 1]. \tag{8}$$

The size of $f_{\max} - f_{avg}$ is often used to measure the degree of evolution of a population. The method is indeed valid. However, the shortcoming is that the fitness function is designed based on a specific problem and its value varies with the specific problem. Moreover, during the calculation of the algorithm, the maximum and minimum fitness of the population also change continuously, so it is difficult to determine the appropriate threshold to judge the evolutionary degree of the population. The smaller the value of E_1, the closer the average fitness of the population to the optimal fitness of the population, which means that the population is more evolved. On the contrary, it means that the population is less evolved. Formula (9) indicates the variability of the population individuals.

$$E_2 = \frac{1}{T}\sum_{x=1}^{T} \frac{f(i_x) - f_{\min}}{f_{\max} - f_{\min}} \in [0, 1]. \tag{9}$$

The smaller the value of E_2, the more dispersed the population adaptation, the greater the population variation, and the better the diversity, and vice versa, the poorer the population diversity.

In formulae (8) and (9), $f(i_x)$ denotes the fitness of individual i_x. $f_{\max}$ is the maximum value of population fitness. $f_{\min}$ is the minimum value of population fitness. f_{avg} is the average value of population fitness. T is the number of population chromosomes.

Crossover arithmetic is used to obtain superior individuals by recombinant development of existing genes. If all individuals in the population do not have a certain gene, the missing gene cannot be obtained by crossover arithmetic in any way. If the population only generates new genes without further exploration, the good individuals in the population will be destroyed continuously, and convergence will be delayed. When the population is in a stagnant state, the population convergence is higher if the population is more concentrated in terms of adaptation. This will lead to more similarity between chromosomes and poor population diversity. In this case, it is necessary to replace the evolutionary strategy that can generate new genes to expand the population diversity. If the population fitness is more dispersed, the population convergence is low, the similarity between chromosomes is low, and the genes of chromosomes are more abundant in the population. At this time, the main focus should be on exploring the solution space of the current individual and the need to replace the evolutionary strategy that can exploit the existing genes.

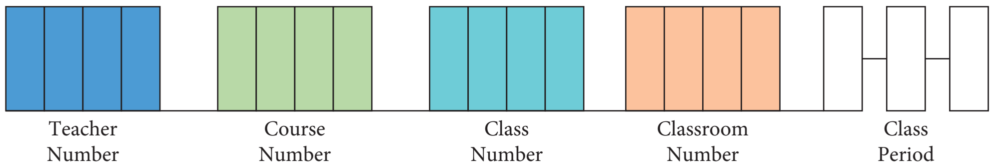


Figure 4: Example diagram of gene structure.

3.3. Multistrategy Ant Colony Algorithm

3.3.1. Initial Pheromone Nonuniform Distribution Strategy. In the solution process, one of the main bases for path selection by ants is the pheromone content of intergrid paths. In the original ant colony algorithm, the initial pheromone of each intergrid path is uniformly distributed and is a constant value. The value of the heuristic function of the ants will determine the magnitude of the state transfer probability, which is known by its calculation method. The difference in the heuristic function values of different grids is small, and without pheromone guidance, the blindness of the ant search is large, and the quality of the solution obtained from the search is low. Nonuniformization of the initial pheromone of the intergrid path is beneficial to accelerate the convergence speed of the algorithm. The heuristic function idea of A^* algorithm is used to differentially process the pheromones of different interraster paths. The corresponding pheromone concentrations are assigned according to their coordinates, and the calculation process is as follows:

$$\begin{aligned}\tau_{xy}(0) &= \tau_0 + \Delta\tau_{xy},\\ \Delta\tau_{xy} &= c\left(\frac{1}{f(x)}\right)\cdot\left(\frac{1}{f(y)}\right),\\ f(i) &= \varepsilon\cdot a(i) + (1-\varepsilon)\cdot b(i),\end{aligned} \tag{10}$$

where τ_0 is the original initial pheromone value. $\Delta\tau_{xy}$ is the additional pheromone added to the (x, y) path. c is a constant value, taken according to the empirical value and the raster scale. ε is the weighted value. $a(i)$ is the spacing between the starting point and the current point. $b(i)$ is the spacing between the current point and the end point.

Based on the shortest distance between two points, the closer the ant's path is to the starting and ending line, the better the path is. When the next grid to be selected is closer to the end point, the path is better. Therefore, ε should be taken as a small value so that the end point plays a dominant role. Besides, for the overall effect of the algorithm, if the gap between the nonuniform pheromone values is too large, it will cause the algorithm to converge too early and lead to local optimum problems, and if it is too small, the strategy will have limited effect. Therefore, the values of c and ε were experimentally compared several times, and the final values were determined as $c = 2\times4\times4 = 32$, $\varepsilon = 0.1$.

3.3.2. Directional Neighborhood Expansion Strategy. In the original raster map-based algorithm, ants can only select the next path node in four or eight neighboring raster grids around them. The search direction is limited, and the optimal path length found due to the step length limitation is also long. The idea of this extended neighborhood is introduced, and the neighborhood is extended directionally according to the relative positions of the starting and ending points of the map, as shown in Figure 2.

The solid line represents the feasible raster of the original ant colony algorithm. The dashed line represents the feasible grid added after the directed neighborhood expansion. Using the directed neighborhood expansion strategy not only enriches the search direction of ants and achieves the purpose of finding shorter paths through one search, but also reduces the computational effort of the algorithm and speeds up the operation of the algorithm. Due to the increase of the searchable grid range of ants in the improved algorithm, the search method of ants is reset. In the eight surrounding grids of the current grid, ants may not select an obstacle grid (hard constraint) as the next move path node in the extended neighborhood grid.

3.4. Genetic Ant Colony Hybrid Optimization Algorithm Process.

The improved genetic ant colony algorithm flow is given in Figure 3. By using the ant colony algorithm, a better set of solutions is generated after one iteration, which is used as the initial population of the genetic algorithm. This is an effective way to reduce the number of times the genetic algorithm seeks to find the optimal course scheduling result quickly.

4. Course Scheduling Implementation

4.1. Gene Coding and Chromosome Construction

(1) The teacher number, course number, class number, classroom number, and class period form a tuple, which is the gene of the genetic algorithm. Each gene can be regarded as a classroom unit in the class schedule, and the gene structure is shown in Figure 4. For example, in gene code 1001_2001_3001_4001_0-1-0, it means that teacher 1001 is teaching class 3001 in classroom 4001 during lesson period 0-1-0. The course number is 2001. u_x = (day, division, section). The day indicates the day of the week; division indicates the hour (AM, PM); section indicates the class period.

In practice, the instructor needs to set up the schedule for each semester before scheduling; i.e., the relationship between courses, classes, teachers, and classrooms is determined.

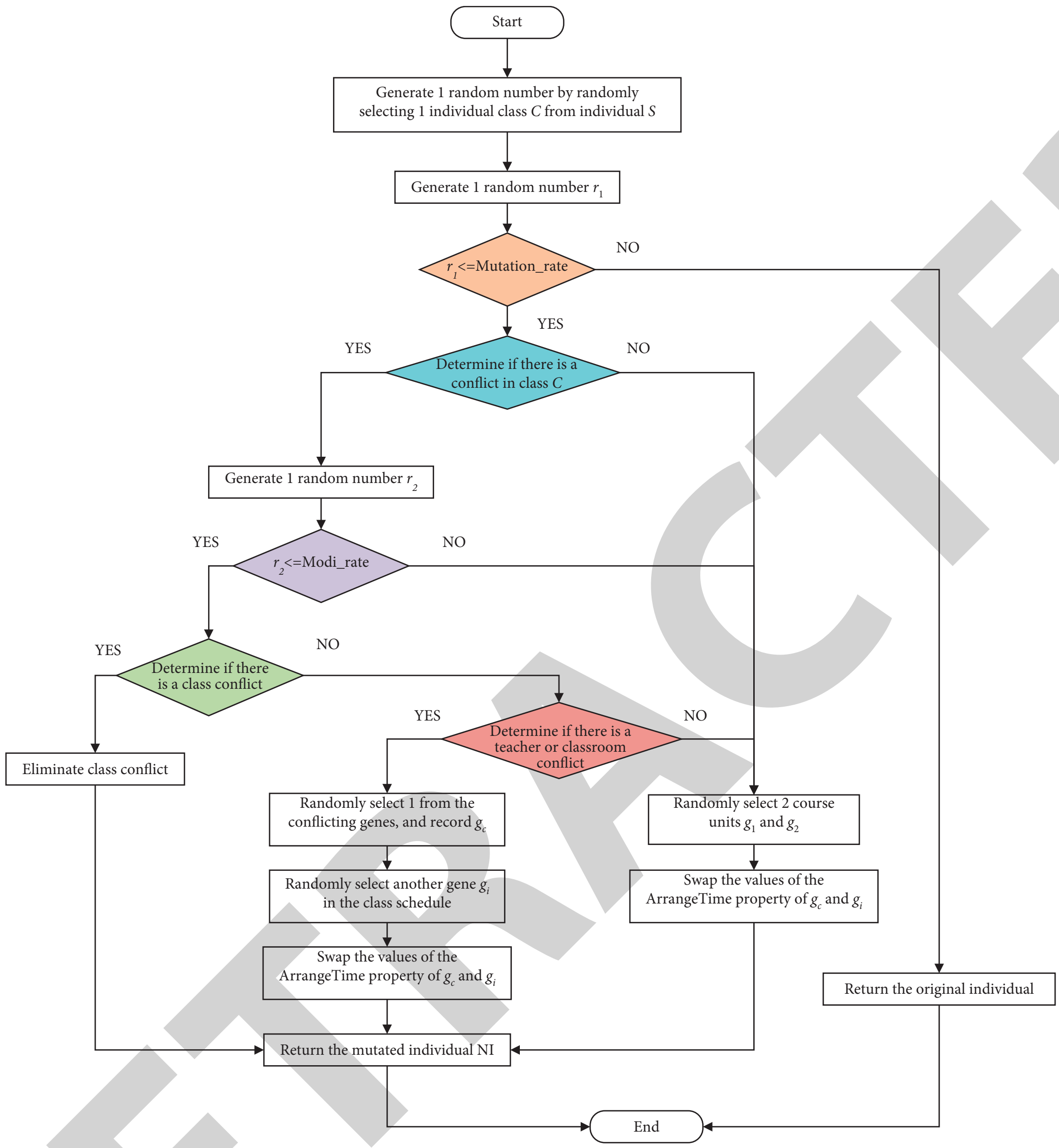


Figure 5: Flowchart of mutation operation.

(2) Construction of chromosomes on the basis of gene coding. The class schedule units (genes) of all classes form 1 chromosome (individual). The CourseInfo object represents the course information of each course in the schedule, including the course, class, instructor, classroom, campus, and weekly class hours. Each CourseInfo corresponds to one or more genes. This is determined by the number of weekly sessions under the CourseInfo. Each individual has the same number of genes.

4.2. Adaptation Function. In this paper, the fitness function is considered in terms of the priority and uniformity of the course sessions, and punishment degree is added to reduce the violation of constraints.

In summary, the fitness function in this paper is given by

$$f = (\omega_1 \cdot f_1 + \omega_2 \cdot f_2) \cdot \beta, \qquad (11)$$

where ω_1 and ω_2 are the weights, and $\omega_1 + \omega_2 = 1$. $\beta = \alpha_1^{\text{hard\_}vios} \cdot \alpha_2^{\text{soft\_}vios}$ is punishment degree. α_1 and α_2 are the penalty factor, $\alpha_1 < 1$, $\alpha_2 < 1$, and $\alpha_1 < \alpha_2$.

4.3. Design of Crossover and Mutation Operations in Scheduling System. The crossover strategy in this paper is as follows: individuals Individual_1 and Individual_2 are selected by the roulette algorithm based on individual adaptation. 1 grade is randomly selected, and the set of gene objects is collated. Exchange the gene objects of the classes involved in these 2 individuals.

The idea of gene editing is introduced in the mutation operation in this paper.

Step 1: 1 class is randomly selected and all gene objects related to the class are obtained from the individuals to be mutated.

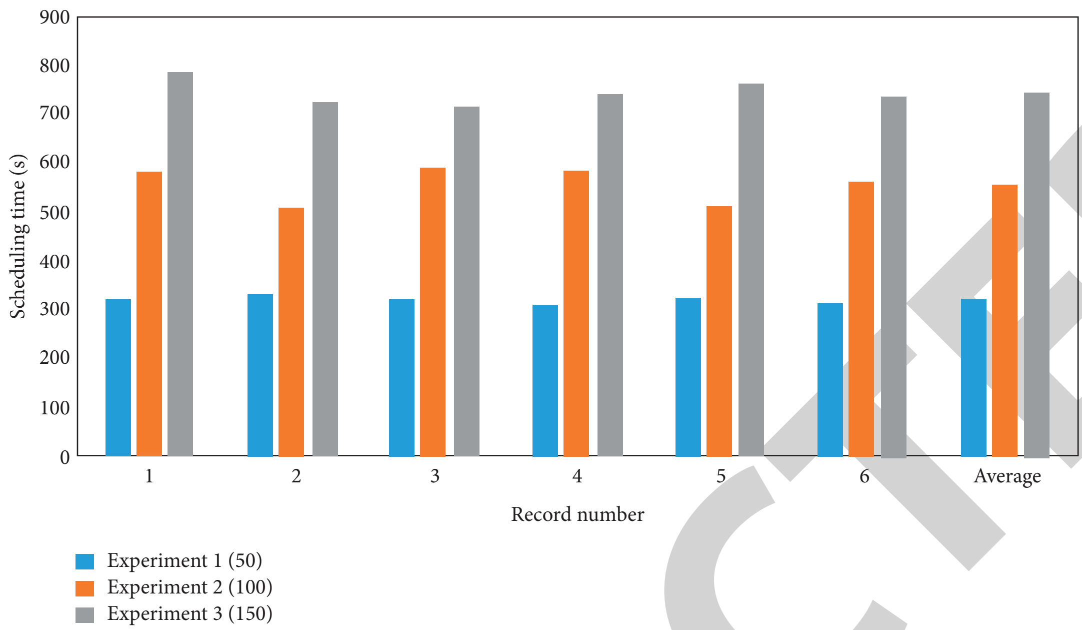


Figure 6: Effect of population size on scheduling time.

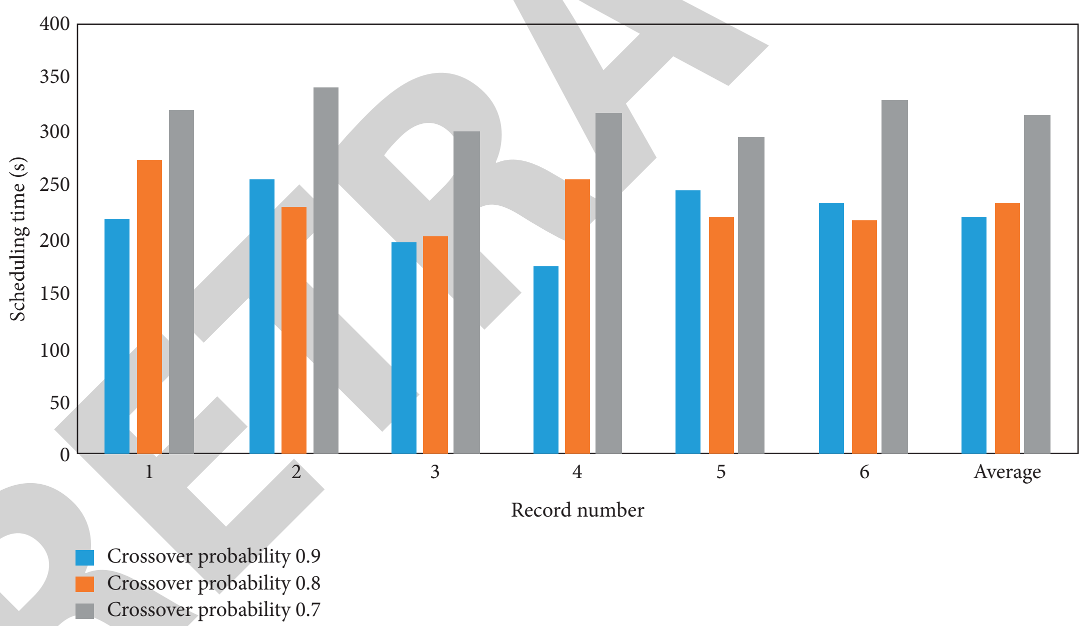


Figure 7: Effect of crossover probability on scheduling time.

Step 2: determine whether the set of gene objects has conflicts. If the constraint is violated, step 3 is executed; otherwise step 5 is executed directly.

Step 3: identify if the course schedule has a conflict (2 or more courses are scheduled for the same class period). If there is, only 1 class is reserved for that period, and the rest of the classes are scheduled for the free period. That is, select $(i-1)$ gene objects from the i conflicting gene objects and set their arrangeTime property value to the unscheduled free period, ignoring whether it will cause other conflicts. If there is no class conflict, then step 4 is executed; otherwise, the current mutation operation is exited.

Step 4: identify if the teachers and classrooms schedule has a conflict. If yes, swap the conflicting course schedule with another schedule. After step 4, save the mutated individuals and exit the mutation operation.

Step 5 : randomly select 2 courses and swap them.

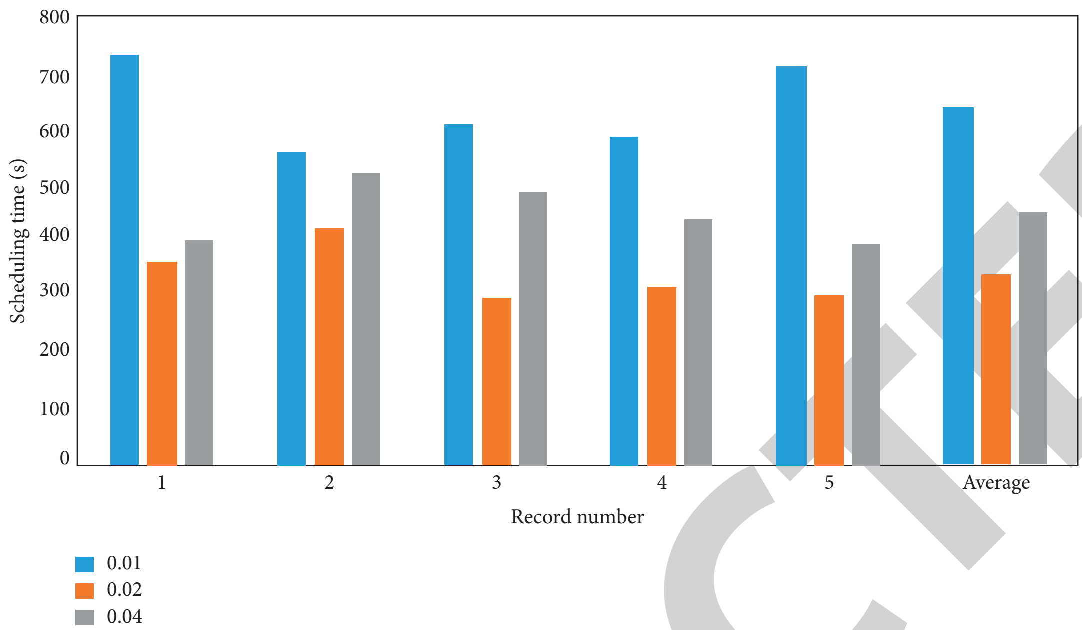


FIGURE 8: Effect of mutation probability on scheduling.

The overall flowchart of each mutation operation is shown in Figure 5, where Mutation_rate is the mutation probability and Modi_rate is the genetic modification probability.

4.4. Termination Conditions. When the class size is large, the individuals obtained from 1,000 iterations of the population may still have a certain gap with the optimal solution, so the maximum number of genetic iterations should be automatically adjusted according to the class size. The termination conditions of evolution are given:

(1) The population is iterated to $125 \times$ classNum times, where classNum denotes the number of classes in the school.

(2) The best individual in the population is conflict-free, and the ratio of the average fitness of the individuals in the population to the fitness of the best individual is greater than 0.9.

(3) The best individual in the population is maintained continuously for $5 \times$ classNum generations.

5. Result Analysis and Discussion

5.1. Parameter Setting. The university scheduling algorithm based on improved genetic ant colony hybrid optimization was performed in a microcomputer environment with 2 GB RAM, Core 2 Duo2.2 GHz CPU, Windows 7, Visual C++2008. The proposed algorithm was tested using data from the School of Mathematics and Information Systems of a university.

Time slots: Classes are held 5 days a week, with 5 classes scheduled each day, for a total of 25 slots.

Classrooms: 56 (including all scheduled classrooms and computer rooms).

Practical training room: 25.

Teachers (including external teachers): 180.

Classes: 63 (maximum number of students per class is 50).

Students: 3620.

According to the principle of genetic algorithm to search for optimal solutions, the following factors were considered experimentally.

First, the effect of population size on scheduling: the main reference is the time required to find the optimal solution. How large the population size is moderate, so that it is more cost-effective for scheduling. Therefore, it is important to find a moderate population size.

Second, the impact of crossover probability on scheduling: the size of the crossover probability affects the convergence of the search for the optimal solution; the larger the value the faster the convergence, but too large will converge too early for a better global search. Therefore, it is important to find a moderate value of crossover probability.

Third, the impact of variance probability of scheduling: too large variance probability will cause the genetic algorithm to be unstable. Too small variation probability will in turn make the global search difficult. Therefore, it is important to find a moderate value of variation probability.

Fourth is the number of generations of genetic evolution at convergence.

5.2. Experimental Comparison. The effect of population size on scheduling is given in Figure 6. The population size was tentatively set at 50, 100, and 150, and three groups were tested, with six results recorded for each set of experiments.

Table 1: Comparison of metric performance of several algorithms.

Algorithm	Conflict rate (%)	Success rate (%)	Run time (s)	Scalable soft constraint
Proposed algorithms	0	100	36	YES
Greedy algorithm	44	57	94	NO
Genetic algorithm	29	72	164	NO
Ant colony algorithm	18	85	127	NO
Genetic ant colony algorithm	13	91	113	NO
Annealing algorithm	52	49	234	NO

Table 2: Comparison of algorithm adaptability.

Scheduling unit	Genetic algorithm	Ant colony algorithm	Proposed algorithm
66	178.9	197.6	210.4
124	200.6	219.8	227.5
239	229.5	247.5	269.7
528	268.6	283.9	319.6
849	297.4	317.8	353.8

Obviously, in agreement with the common knowledge and prediction of this problem in this paper, the smaller the population size, the shorter the time to find a suitable solution. However, the population size is too small to find the optimal solution within a limited number of genetic generations. From the experiment in Figure 6, it can be seen that population size 100 is a more moderate result.

The effect of crossover probability on scheduling is given in Figure 7. In this paper, experiments were conducted for crossover probabilities of 0.7, 0.8, and 0.9.

From the above experimental data in Figure 7, it can see that the larger the crossover probability is, the less time it takes to find a joint solution, which is actually the faster the search process converges. However, too large will converge too early to perform a better global search. Therefore, in this paper, the crossover probability is taken as 0.8, which is a more moderate result.

The effect of mutation probability on scheduling is given in Figure 8. If the mutation probability is too high, the algorithm will be unstable. According to genetic knowledge, too much variation does not produce good individuals but makes each chromosome unstable and produces a lot of conflicting results. Mutation probability is too small, too few new individuals are generated, and global search is difficult to carry out. In this paper, the probability of variation is 0.01, 0.02, and 0.04. From the experimental results in Figure 8, the difference between the experimental results of different mutation probability can be seen. The mutation probability of this system is 0.02, which is a moderate result.

Based on the original scheduling data, experiments were conducted for different penalty factors parameters. It can be obtained that when the probability of variation is 0.02, the penalty factor α_1 is 0.88, and the penalty factor α_2 is 0.94, and the fitness of the final result is the highest, reaching 0.83729.

5.3. Other Performance Analyses. Experimental simulations were conducted to compare with the greedy algorithm, genetic algorithm, ant colony algorithm, genetic ant colony algorithm, annealing algorithm, and the proposed algorithm to derive the performance differences of each algorithm in terms of conflict rate, success rate, running time, and scalable soft constraints for the metrics. The algorithm in this paper performs 20 simulation experiments with certain course weights, and the average performance values are shown in Table 1. It can be seen that the performance of the scheduling algorithm in this paper is better.

Different scheduling units such as 66, 124, 239, 528, and 849 were set and the adaptation values were compared under the above different scheduling units to obtain the results in Table 2.

The results in Table 2 show that the fitness values of the three algorithms increase simultaneously with the increase in the number of scheduling units. After the number of scheduling units reaches a certain level, the fitness of the three algorithms tends to stabilize. Keeping the number of scheduling units constant, the fitness value of the hybrid algorithm is higher than that of the single algorithm, which indicates that the hybrid algorithm is feasible and effective in the application.

6. Conclusion

The structural design and functional perfection of the scheduling system are increasingly demanding in view of the trend of diversity in university curriculum. Administrators should study the theory and practice of scheduling system design from the practical operational level. More intelligent, effective, and feasible algorithms are continuously explored to solve the problems encountered in the process of system development and actual operation. This paper presents a college scheduling algorithm based on improved genetic ant colony hybrid optimization. The algorithm proposes an enhanced infection genetic algorithm, a directed neighborhood expansion strategy, and a penalty mechanism and innovatively introduces the idea of gene editing in the genetic operator of the genetic algorithm. All these enhance the scheduling performance. This paper is based on the actual needs of universities to solve the scheduling problem, and the innovation points are mainly in the following four aspects:

(1) The enhanced infection genetic algorithm is proposed, which removes the unnecessary replication operation in the traditional genetic algorithm and improves the convergence speed, convergence accuracy, and operation speed of the algorithm.
(2) The ant movement rules are redefined using a directed neighborhood expansion strategy in the ant colony optimization algorithm to further shorten the path and improve the search efficiency.
(3) The genetic operator in the genetic algorithm innovatively introduces the idea of gene editing to automatically locate a conflict in the class schedule with a certain probability and eliminate the conflict, which greatly improves the evolutionary effect of the genetic algorithm.
(4) For the initial population generation strategy of the genetic algorithm proposed in this paper, a penalty degree design is introduced into the fitness function, which also improves the evolutionary effect of the genetic algorithm to a certain extent.

Experimental results show that the algorithm proposed in this paper outperforms other algorithms in terms of scheduling conflict rate, success rate, running time, and fitness value. Therefore, the general framework and ideas of the improved genetic ant colony hybrid optimization algorithm for solving the scheduling problem proposed lead to an effective solution to the colleges and university scheduling problem. The future work is to improve the college scheduling algorithm to further improve the efficiency of scheduling.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the Hunan University of Information Technology and Hunan First Normal University.

References

[1] W. D. Pahlawanti, E. Harapan, and D. Wardiah, “The influence of school principal supervision and school committee participation on the quality of junior high school education,” *International Journal of Progressive Sciences and Technologies*, vol. 23, no. 1, pp. 324–333, 2020.
[2] X. Xing, M. Huerta, and T. Garza, “College and career preparation activities and their influence on post-high school education and work attainment,” *Journal of Career and Technical Education*, vol. 34, no. 1, p. 8, 2019.
[3] L. Williams, M. Martinasek, K. Carone, and S. Sanders, “High school students’ perceptions of traditional and online health and physical education courses,” *Journal of School Health*, vol. 90, no. 3, pp. 234–244, 2020.
[4] H. L. Huang, G. J. Hwang, and C. Y. Chang, “Learning to be a writer: a spherical video-based virtual reality approach to supporting descriptive article writing in high school Chinese courses,” *British Journal of Educational Technology*, vol. 51, no. 4, pp. 1386–1405, 2020.
[5] K. P. Kremer, “Predictors of college success outcomes in emerging adults: the role of high school dual enrollment courses,” *Emerging Adulthood*, vol. 10, no. 1, pp. 188–196, 2022.
[6] Y. V. Ermanto and Y. F. Riti, “Comparison of welch-powell and recursive largest first algorithm implementation in course scheduling,” *Journal of Management Science (JMAS)*, vol. 5, no. 1, pp. 05–12, 2022.
[7] A. K. Nugroho, I. Permadi, and A. R. Yasifa, “Optimizing course scheduling faculty of engineering unsoed using genetic algorithms,” *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 7, no. 2, pp. 91–98, 2022.
[8] R. Parks, H. Ajjan, A. Gaebel, and A. Taylor, “Room scheduling: a dependent variable to reduce the spread of COVID-19,” *College & University*, vol. 96, no. 2, pp. 39–41, 2021.
[9] X. Wang, Y. Chen, and X. Wang, “Design and application of experiment teaching arrangement system,” *Technology Wind*, no. 6, pp. 69–71, 2022.
[10] Z. Zaeniah and S. Salman, “Designing class schedule information system by using taboo-search method,” *Pilar Nusa Mandiri: Journal of Computing and Information System*, vol. 16, no. 2, pp. 241–248, 2020.
[11] R. Baker, B. Evans, Q. Li, and B. Cung, “Does inducing students to schedule lecture watching in online classes improve their academic performance? An experimental analysis of a time management intervention,” *Research in Higher Education*, vol. 60, no. 4, pp. 521–552, 2019.
[12] M. Tomić and D. Urošević, “A heuristic approach in solving the optimal seating chart problem,” in *Proceedings of the International Conference on Mathematical Optimization Theory and Operations Research*, pp. 271–283, Springer, Irkutsk, Russia, September 2021.
[13] N. G. A. H. Saptarini, P. I. Ciptayani, and I. B. I. Purnama, “A custom-based crossover technique in genetic algorithm for course scheduling problem,” *TEM Journal*, vol. 9, no. 1, pp. 386–392, 2020.
[14] P. Wang, X. Xu, and C. Liu, “An improved adaptive genetic algorithm and its application in intelligent course scheduling system,” in *Proceedings of the 2019 6th International Conference on Information Science and Control Engineering (ICISCE)*, pp. 121–125, IEEE, Shanghai, China, December 2019.
[15] X. Chen, X. G. Yue, R. Y. M. Li, A. Zhumadillayeva, and R. Liu, “Design and application of an improved genetic algorithm to a class scheduling system,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 01, p. 44, 2021.
[16] S. A. A. Edalatpanah, “A direct model for triangular neutrosophic linear programming,” *International journal of neutrosophic science*, vol. 1, no. 1, pp. 19–28, 2020.
[17] E. Kurtuluş, A. R. Yıldız, S. M. Sait, and S. Bureerat, “A novel hybrid Harris hawks-simulated annealing algorithm and RBF-

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9837091, 1 page
https://doi.org/10.1155/2023/9837091

Retraction

Retracted: Multiobjective Optimization Algorithm for EFRM Strategy

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Jiang, X. Wang, and X. Wu, "Multiobjective Optimization Algorithm for EFRM Strategy," *Security and Communication Networks*, vol. 2022, Article ID 4994383, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4994383, 12 pages
https://doi.org/10.1155/2022/4994383

Research Article

Multiobjective Optimization Algorithm for EFRM Strategy

Weiwei Jiang,[1] Xi Wang,[2] and Xuefeng Wu[1]

[1]*Applied Technology College, Soochow University, Suzhou 215325, Jiangsu Province, China*
[2]*Economic College, Jiaxing University, Jiaxing 314001, China*

Correspondence should be addressed to Xi Wang; wangxi@zjxu.edu.cn

Received 25 February 2022; Revised 29 March 2022; Accepted 26 April 2022; Published 13 June 2022

Academic Editor: Chin-Ling Chen

With the rapid development of the global economy in recent years, market competition has become more and more intense. Therefore, the market competitiveness of the enterprises is becoming higher and higher. Among them, the enterprise financial risk management (EFRM) is one of the key factors that decide the enterprise market competition. It also shows that enhancing the enterprise market competition needs to strengthen the enterprise's financial control and management, with the method of reasonable controlling of enterprise financial risk (EFR). The current research and development status of EFRM is combed, the theory of financial internal control and its significance are expounded, the related concepts of the company's financial internal control are studied, and the definition, classification, basic features, and causes of financial risk are analyzed in this paper. Finally, a company is taken as research object, the characteristics and status quo of the company's financial risk system are studied, and its existing problems are analyzed, a mathematical model for the company's FRM is built, and the multiobjective optimization algorithm is used to optimize the model, to improve the level of a company's FRM performance and market competitiveness and to guard against the company's financial risk.

1. Introduction

With the process of economic globalization, the market economy competition environment is becoming more and more intense; there are many companies bankrupt because of poor management around the world every day. Improving enterprise management model is an effective way to improve the enterprise market competitiveness, which attracted more and more attention from people [1–4].

Figure 1 shows the schematic diagram of enterprise risk classification.

According to Figure 1, it shows that the risk of enterprise can be divided into enterprise financial risk (EFR) and enterprise management risk (EMR). EFR can be divided into funding risk, investment risk, return on risk, and distribution risk. And EMR can be divided into development of failure risk, risk of product quality standardization, evil competition risk, and risk of product out of date. Enterprise risk is everywhere [5–8]; thus, optimizing company management strategy, making enterprise risk in a reasonable evaluation is an important way to keep an enterprise developing healthy for a long time.

Financial risk management (FRM) is a branch of risk management, which is a risk management method developed from experience. Mellichamp et al. [9–11] studied the development of the market economy law, analyzed the concepts such as profit, risk, and investment, and studied the method of optimization to reduce the enterprise operational risk. Lu and Chen [12] took network company as a research object, the convolutional neural network method was used to establish the network company supply chain risk model, the researchers of the supply chain risk characteristics are analyzed in detail, and some targeted suggestions are put forward, in order to reduce the company's operational risk providing a certain theoretical support.

The development of the financial internal control theory mainly experienced the following stages [13–20].

First is the infancy stage of early financial. The ancient Roman palace "double billing system" is the earliest financial internal control system, and the system clearly uses the methods of double bookkeeping, regular assessment, and examination of cheating to achieve the purpose of supervising the economy and controlling financial revenue and expenditure.

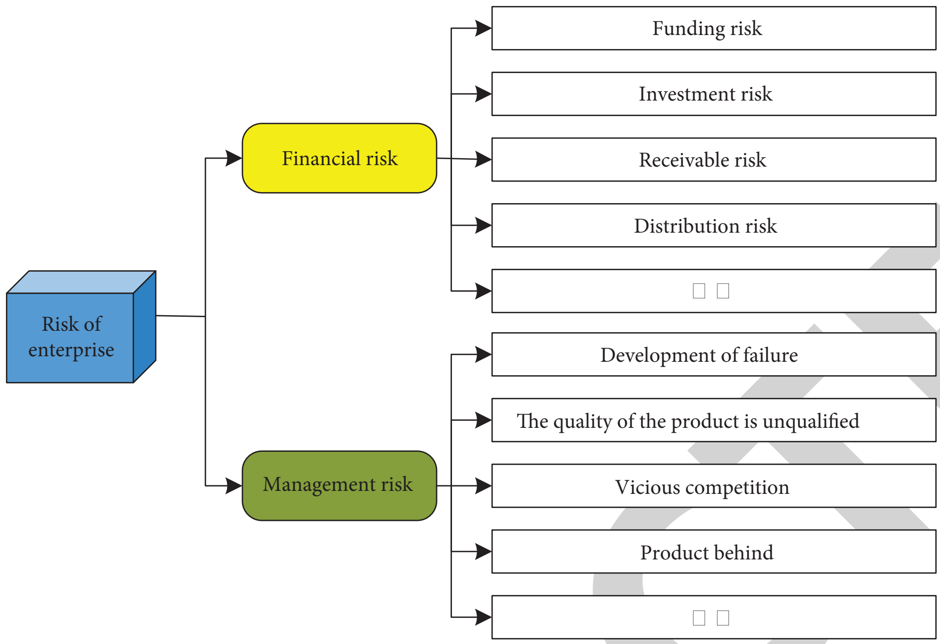


Figure 1: Schemes of enterprise risk classification.

Second is the development stage of the modern financial internal control system. Since the second industrial revolution, the international economic market developed greatly, governments and companies have gradually realized the importance of financial securitytherefore, the research on financial internal control theory and financial internal control mode has been paid more and more attention, and has been widely practiced and promoted in the company's operation. At the beginning of the 20th century, American Lawrence [21–23] created the financial security management method in the first time and put forward the concept of internal check. It mainly refers to financial personnel management, which means the method of clear division of duties, realizing people or departments to manage economic business. And combining with the method of cross check and cross control to restrain each business management, it also has strict accounting registration system to implement the company's internal financial personnel management of the company's financial management. In the middle of the 20th century, there was a fast development and promotion of the concepts and theories of enterprise financial internal control primarily because it is to protect the enterprise assets. Enterprise data accounting is an important mean of accuracy reliability recording, which can improve economic efficiency of enterprises and ensure enterprise policies run smoothly.

Third is the mature stage of financial internal control. There several important landmark events of financial internal control development mature as follows [24–30]. (1) In 1985, the U.S. government organized the national anti false financial reporting Committee and COSO committee jointly with the Institute of certified public accountants, the Accounting Association, the managers' Association, the Institute of internal auditors, the Institute of management accountants, etc. The institute of internal auditors, management accountant association formed the national commission against false financial report and the Committee of sponsoring Organizations (COSOs). They studied the method of avoiding fraud in financial statements. (2) In 1988, American audit committee issued accounting statements audit in the attention to the internal control structure; it was the first time the financial internal control theory was put forward and pointed out that the financial internal control structure is to use various policies and procedures on the enterprise's financial situation and economic goals, which can provide a reasonable and effective guarantee. (3) In 1992, the American COSO committee explained that the company's financial internal controls the composition of the factors in detail once again and pointed out that the control activities, the control environment, risk assessment, information and communication, and supervision constitute are the important factors of internal control. In the same year, the card DE Burleigh company of British took corporate governance as a breakthrough point and studied the quality of financial reporting and the relationship between the financial internal control and corporate governance. The research results showed that financial internal control is one of the frameworks of corporate governance, improving that the internal control mechanism is the clear requirement of corporate governance. These financial internal control typical events marked the financial internal control theory as gradually mature.

Fourth is the new development stage of financial internal control. In 21st century, the United States credibility problem caused exposure of more than a dozen companies such as Enron, WorldCom's company financial fraud [31, 32]. It is seriously influenced the whole bidding market for a long time and the economic recovery of U.S. These series of events showed that enterprise financial internal control institution running failure will bring serious consequences, but it will promote the development of the company's financial internal control mechanism and perfect. At present, the relevant financial internal control system and the rules and regulations have been studied and specified by the government and enterprises, and they played a positive role in assistance and specification.

At present, the financial internal control has become an important mean of enterprise management. It has the advantageous of enterprise resources reasonable configuration, which can improve enterprises management efficiency and helpful to risk prevention, financial controlling, and reducing loss. Related theory of EFRM was introduced in this paper, the model of EFRM was built, and the multiobjective optimization theory and its algorithm were combined to solve the EFRM model. The purpose of this study is to improve the level of EFRM and provide a certain theoretical guidance to enterprise management. Finally, a company is taken as a researching example, the characteristics of the company's FR management structure are analyzed, and the researched EFRM model and its solving method are used to model and solve the model of company's financial internal control system, which will provide a theoretical support to improve the enterprise management level.

2. Related Works

2.1. Introduction of Business Risk-Related Theory. The enterprise risk management process is shown in Figure 2.

Due to Figure 2, the enterprise risk management is mainly composed of risk management records, monitoring, planning, risk response, risk scope definition, and risk orientation (managing risk log and monitoring risk, planning and risk response, define the scope and the identify risks, risk analysis) six function modules. In the process of enterprise risk management, these six functional modules are interrelated and affect each other, and there is no established sequence for disposal.

In this paper, the EFRM strategy optimization problem is studied; the first FR refers to the enterprises in the financial activities because the internal environment is all sorts of unpredictable or uncontrollable factors, in a certain period of enterprise's actual financial income and financial revenue forecast deviation, thus the possibility of loss.

The EFR can be divided into financing risk, investment risk, return on risk, and distribution risk, as shown in Figure 1. In order to optimize the EFRM strategy, it needs to analyze the various indicators of risk and control risk reasonably.

EFR is everywhere, and risk will change with the progress of the environment and the event constantly. Therefore, risk identification, risk measurement, and risk controlling should be changed over time.

Employees should understand risk correctly, grasp the characteristics of risk, insight into the dialectical relationship of risk and environmental changes, and master the methods of risk management. It gives vital importance to the event, and an important matter of enterprise must be careful treatment.

Risk is inevitable in the process of enterprise operation. In terms of efficiency, generally high risk comes with high benefit and low risk corresponds to low income. In the progress of enterprise financial management, financial statements can provide enterprise's financial position, operating results and cash flow, and other financial information to managers, which will provide the foundation for enterprise financial management decision support.

Nowadays, the main measures used to guard against EFR are as follows. First, capital flow: (1) Improving the efficiency of fund using: this is the basis for prevention and controlling of fundraising risk because companies servicing money comes from earnings. If the enterprise is with poor management and long-term losses, although it has an effective cash management, it will lead to the pressure of the enterprise unable to pay the debt principal and interest on schedule. (2) Moderating debt and optimizing the capital structure: indebtedness is a double-edged sword; it brings higher yields but also brings greater funding risk loss. So, enterprises must be moderate debt management. Determining the "degree" of moderate debt is a complex and difficult problem. In theory, application of the optimal capital structure theory can meet the low comprehensive cost of capital and raising enterprise value maximization. In practice, it should be adapted to the specific circumstances of the enterprise. For some enterprises with good production business operation and faster moving capital stock, its debt ratio may be more appropriate. While, for enterprises with poor operation and slow capital turnover, their debt ratio should be appropriately lower. (3) Reasonable collocation of current liabilities and long-term liabilities: the proportion of current liabilities and long-term liabilities shall be corresponded to the enterprise capital condition.

Second is about investing. (1) Strengthening the feasibility study on investment plan: if an enterprise can make reasonable projections for the future earnings before investment, the scheme of high risk and low profits should be excluded. Money should only be put into those feasible solutions, which will prevent and control the investment risk. (2) Application of portfolio theory to portfolio reasonably: based on portfolio theory, if other conditions are constant, the smaller the correlation coefficient of different investment rates of return, the greater the ability of reducing the risk of overall investment portfolio. Therefore, in order to achieve the goal, it should spread investment risk, when making investment decisions must pay attention to the analysis of the correlation between investment projects. In the progress of securities investment, enterprises should purchase different securities to reduce the correlation coefficient of different industries. If the purchase of securities within the same industry, it should avoid all of the purchased securities from the same company.

Third is about funds recovery. (1) Choosing reasonable sales way and the payment method: for the customers with good financial and credit conditions, taking the way of sell on credit and to control the total credit amount within a line of credit, settlement is taken the instalment payment and commercial drafts when the less risky way of settlement. While, for those customers with poor credit condition, the solvency of customers should adopt the mode of the pin, taking corresponding methods of payment of exchanging,

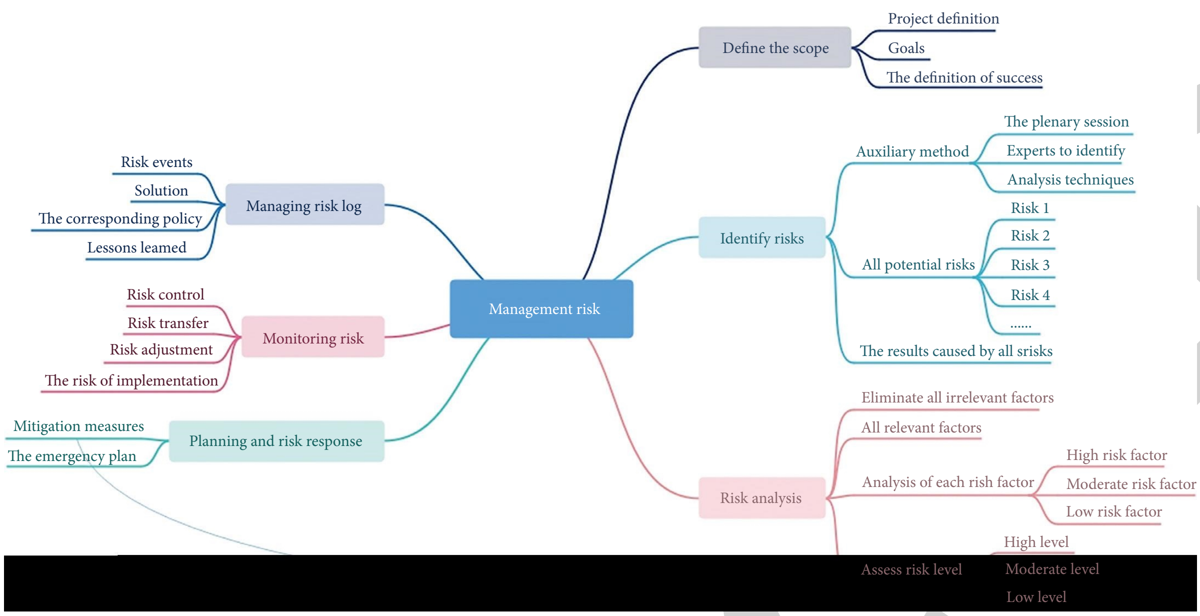


Figure 2: Diagram of enterprise risk management process diagram.

checking as far as possible. (2) Establishing the reasonable collection policy and collecting payment for goods timely: to customers overdue outstanding credit arrears, the enterprise should organize personnel to collection. (3) Establishing a system of bad debts reserves: The enterprise shall, in accordance with the principle of prudence, withdraw bad debt reserves for possible bad debt losses before they occur, so as to reduce the falsely increased profits in the current period and prevent the adverse effects of capital recovery risks.

Fourth is about distribution of income distribution. (1) Making reasonable income distribution policy: the income distribution policy of the company depends on the actual situation of the company profit. Co., LTD.; for example, if a company's surplus is stable, it can extend higher dividends; otherwise, it can only extend lower dividends. (2) Setting up enterprise good image and building investor confidence: income distribution policy, improper or frequent changes in income distribution policy may cause adverse effect to the enterprise. Enterprises should take measures to convey the positive beneficial information to investors actively. Particularly noteworthy is that enterprise cannot disclose false information to restore investor confidence. Otherwise, it will not only unfavorable to the promotion of enterprise value but also it will increase the EFR.

2.2. Multiobjective Optimization Theory. At present, the main classification of multiobjective optimization theory is (1) the optimization of unconstrained and constrained conditions; (2) the uncertainty and randomness of the optimal problem (variable); (3) the linear and nonlinear optimization (that is, the objective function and constraint conditions whether linear); and (4) static and dynamic planning whether changes over time (optimization).

Usually, the multiobjective optimization problem is composed of multiple objective function and some related equality and inequality constraints; mathematical expressions can be described as formulas (1) and (2):

$$\begin{cases} \min f_1(x_1, x_2, \ldots, x_n), \\ \vdots \\ \min f_r(x_1, x_2, \ldots, x_n), \\ \max f_{r+1}(x_1, x_2, \ldots, x_n), \\ \vdots \\ \max f_m(x_1, x_2, \ldots, x_n), \end{cases} \tag{1}$$

$$\begin{cases} g_i(x_1, x_2, \ldots, x_n) = 0, & i = 1, 2, \ldots p, \\ h_j(x_1, x_2, \ldots, x_n) > 0, & j = 1, 2, \ldots q, \\ x_k \in R, & k = 1, 2, \ldots n. \end{cases} \tag{2}$$

$f_1(x_1, x_2, \ldots, x_n), \cdots, f_m(x_1, x_2, \ldots, x_n)$ in formula (1) are called the m target functions, $x_1, x_2, \ldots, x_n$ are called the n variables, and g_i and h_j are equality and inequality constraint condition, respectively.

Combining formulas (1) and (2),

$$\begin{cases} \min & f(x) = [\, f_1(x) \;\; f_2(x) \;\; \cdots \;\; f_m(x)\,]^T \\ & x \in X \quad X \in R^n \\ \text{s.t.} & g_i = 0 \quad i = 1, 2, \ldots, p \\ & h_j = 0 \quad j = 1, 2, \ldots, q, \end{cases} \tag{3}$$

where min $f(x)$ is minimizing vector *f*(*x*), i.e., the vector of the target function is as far as possible to minimization. $X \in R^n$ is a variable constraint set.

Multiobjective optimization problem is about letting multiple targets achieve the best possible condition in a certain range of area at the same time. The solution of multiobjective optimization is a set of equilibrium solution usually. Among them, the non inferior solution means that there is no optimal solution for the multiobjective optimization problem, and all possible solutions are called non inferior solutions, also known as Pareto solutions. However, all of the possible solutions are known as the Pareto solutions. Usually, there is a mutual coupling relationship between multiple targets, may be a plenty of positive correlation, or a plenty of negative correlation (namely, a target optimization, another deterioration). So, in progress of solving multiobjective optimization problems, it needs to balance the goal of the relationship among the find the key goals.

Multiobjective optimization problems usually do not exist the only global optimal solution. In the process of multiobjective optimization, dominant (Pareto dominate) and the optimal (Pareto optimal) concepts are used widespread. Supposing two decision vectors a, $b \in X$, Pareto solution a is better than b, and $a > b$. According to formal (1), it can be described as follows:

$$\begin{cases} f_i(a) \leq f_i(b), & i = 1, 2, \ldots, r, \\ f_i(a) > f_i(b), & i = r+1, r+2, \ldots, m. \end{cases} \quad (4)$$

At present, the main algorithms for solving the above multiobjective optimization problems are linear programming and genetic algorithm. Generally, the objective function is not as linear function, so in this paper, the important research multiobjective optimization genetic algorithm is used to solve the EFRM strategy optimization problem. The multiobjective genetic algorithm is used for analysis and an evolutionary algorithm is for solving the problem of multiobjective optimization; its core is to coordinate the relationship between each objective function and find out the optimal solution set that makes each objective function reach the larger (or smaller) function value as much as possible.

Genetic algorithm in solving multiobjective optimization problem has properties of fast convergence speed and can quickly solve multiobjective optimization problems of pareto solutions and so on. So, the genetic algorithm is used to solve the enterprise financial risk management multiobjective optimization strategy in this article. The genetic algorithm flow chart is as shown in Figure 3.

According to Figure 3, it shows that the process of application of the genetic algorithm (GA) to solve multiobjective optimization enterprise financial risk management strategy is as follows: (1) according to the characteristics and the scope of the independent variables to solve the model, set the target number, fitness, and the parameters such as number of iterations; (2) Set discrete variables according to population parameters; (3) set (2) in the generation of discrete variable parameters into the model of objective function to solve the calculation; (4) The parent population is updated, crossovered, and mutated to produce the next generation population. And repeat steps (1) ~ (3) until convergence, obtain the optimal solution, and exit the multiobjective optimization calculation.

2.3. EFRM Model. Based on the EFRM theory introduced in Section 2.1, EFRM mainly includes funds and employee management. The causes of EFR are excessive borrowing and failed to pay. It showed that the enterprise with moderate debt can enhance its market competitiveness from history experience and has a certain role in promoting enterprise's profit sustained growth ability. Therefore, it cannot simply take corporate debt index to assess the EFR, profitability, and management efficiency.

In order to evaluate EFR scientifically, it needs to set up a scientific and reasonable EFR evaluation model. By integrating a series of financial indicators such as an enterprise's solvency, profitability, and liquidity from different perspectives, we can comprehensively, systematically, and comprehensively analyze and evaluate the enterprise's financial status and operation, which will provide a certain basis for FR management decisions making. According to the actual values of the indexes of EFR, their own satisfaction value and not satisfied efficacy coefficient are calculated. Efficacy coefficient as the index value in the industry level reflects the risk degree of the EFR index in the industry. EFR indicators efficacy coefficient is expressed by the following formula:

$$y_i = \frac{x_i - x_{il}}{x_{ih} - x_{il}}, \quad (5)$$

where y_i is the ith risk indexes of efficacy coefficient; x_i is the ith a risk index actual value; x_{il} to the ith a risk index in the industry is not satisfied with value; x_{ih} is satisfaction value for the ith indicators of risk.

The overall financial risk level of the enterprise is evaluated by the efficiency coefficient of various financial risk indicators of the enterprise. It is assumed that 60 points is the pass line of the efficiency coefficient; formula (6) can be used to calculate the effect of each single index coefficient calculation of their respective index.

$$f_i = y_i \times 40 + 60. \quad (6)$$

The overall FR of the enterprise level evaluation as shown in the following formula:

$$\mathbf{Z} = \sum k_i f_i, \quad (7)$$

where $\mathbf{Z}$ is the EFR level; f_i according to the EFR characteristics determines the ith a risk index (objective function); K_i is the ith risk weights; it is determined by using the fuzzy hierarchy method.

By the clear enterprise risk index f_i, refer to the scale method (in Table 1), which is proposed by United States professor T. L. Saaty, build as shown in formula (8) of the judgment matrix:

$$\mathbf{A}_{m\times m} = \left(a_{ij}\right)_{m\times m} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ & & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}, \quad (8)$$

where a_{ij} is the ith index compared with the jth index of importance. The relationship between a_{ij} and a_{ji} is as shown in the following formula:

$$a_{ji} = \frac{1}{a_{ij}}. \quad (9)$$

For the index weight calculation, firstly the ith line element of the judgment matrix is multiplied, recorded as p_i:

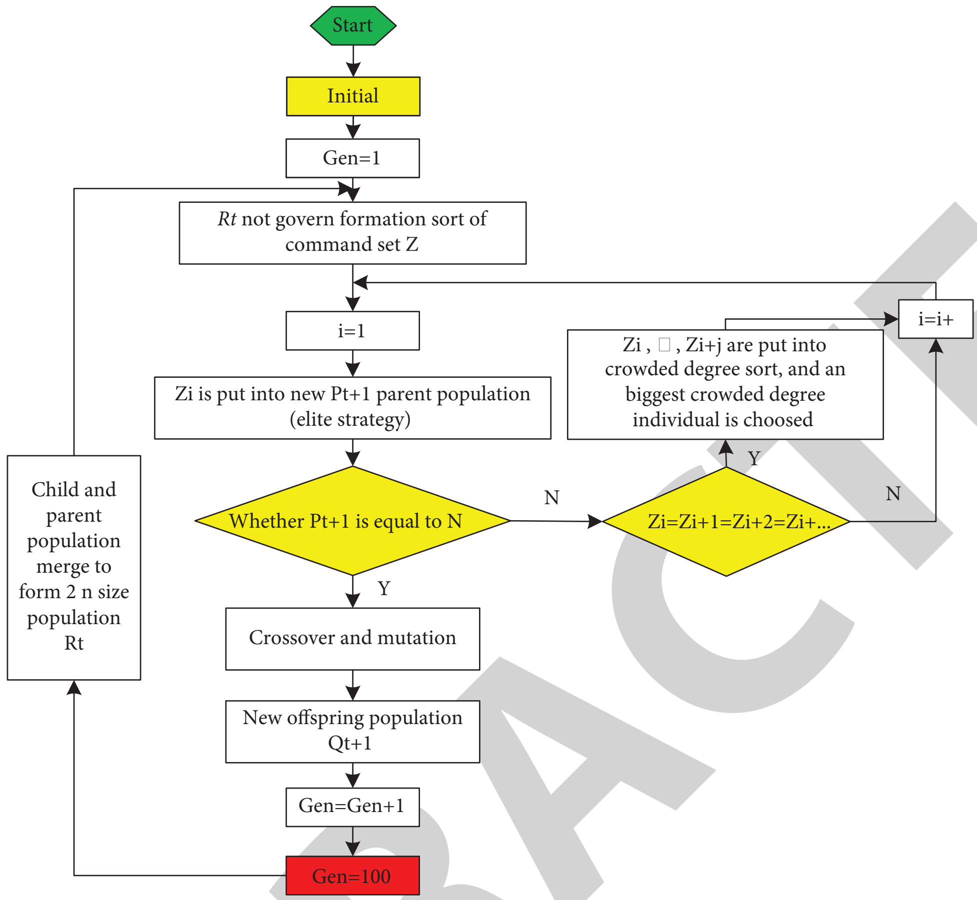


Figure 3: The flow chart of the genetic algorithm.

Table 1: The judgment matrix scale method.

Scale	Definition
1	Two factors are equally important
3	The *i*th index is slightly more important than the *j*th index
5	The *i*th index is more important than the *j*th index
7	The *i*th index is obviously more important than the *j*th index
9	The *i*th index is extremely more important than the *j*th index
2, 4, 6, 8	Between the two adjacent judgment values

$$p_i = \prod_{i=1}^{m} a_{ij}. \tag{10}$$

Secondly, find the m-th root k_{mi} of

$$k_{mi} = \sqrt[m]{p_i}. \tag{11}$$

Finally, the k_{mi} is normalized

$$k_i = \frac{k_{mi}}{\sum_{i=1}^{m} k_{mi}}. \tag{12}$$

Combination formulas (6)–(12) can be obtained by the enterprise's FR values; in combination with the multi-objective optimization algorithm as shown in Figure 3, calculation process can be obtained by the EFR strategy multiobjective optimization scheme.

3. EFRM Strategy Optimization Case Study

3.1. Introduction for Target Enterprise. The company was established 20 years before; it has developed into a given priority with the equipment manufacturing industry. Its business involves the engineering machinery, lifting machinery, coal mining equipment, industrial vehicles, hydraulic components, aviation equipment, military industrial machinery, and so on nearly ten several areas. The level of machine, electricity, liquid integration, and automation control technology of this company is in a leading stage. Robots, unmanned aerial vehicles, and high-end smelting equipment produced by the company are widely used for the company independent development of sensor, controller, and components. At present, the productions of the company have high performance and high-quality engineering machinery, which enjoy high status in the industry. The staffs' number of the company is about 4000, and it is the top 100 global engineering machinery enterprises and the world's top 50 excavator enterprises.

3.2. An EFR Modeling. According to financial statement of the company in recent years, the main financial risk target indexes of the enterprise can be obtained. Based on Section 3.1, the four primary financial risk indexes of solvency, profitability, profit distribution, and financial management ability are used to evaluate the ability of EFRM. Solvency can be divided into cash flow ratio, asset-liability ratio, and the multiple of interest safeguard three secondary indicator. Profitability can be divided into the return on net assets, operating profit margin, surplus cash cover, and cost efficiency. Profit allocation can be divided into research and development of failure risk expenses, staff wage growth, performance reward ratio, and equity incentive ratio. Financial management ability can be divided into financial error ratio, financial personnel quality, and financial personnel in the proportion of members of the company.

The weight relationships among these main financial risk indexes and their comprehensive scores can be obtained by consulting the executives and financial personnel.

Among first-level indicators, in profitability, solvency is as important as the obvious important profits' allocation and is extremely important in the financial management ability. Profit distribution is very important from the financial management ability. As shown in Table 1, primary index important degree of judgment matrix $\boldsymbol{A}$ is as follows:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 7 & 9 \\ 1 & 1 & 7 & 9 \\ \frac{1}{7} & \frac{1}{7} & 1 & 7 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{7} & 1 \end{bmatrix}. \tag{13}$$

Analytic hierarchy process (AHP) is an important tool of the system analysis method about taking the research object as a system, according to the decomposition, comparative judgment, comprehensive way of thinking, to make decisions, and be developed after mechanism analysis and statistical analysis. It has the characteristics such as concise and practical quantitative data and required less, very suitable for the multiobjective optimization model of enterprise financial risk management strategy, so the method used in this paper is to assess and solve the enterprise financial risk management strategy to multiobjective optimization problem.

According to Section 2.3, the analytic hierarchy process (AHP) is introduced to calculate the weights of the four primary financial risk indexes in this paper. They are 0.4314, 0.4314, 0.1044, and 0.0328, respectively, so the weight vector **a** can be recorded as follows:

$$\mathbf{a} = \begin{bmatrix} 0.4314 \\ 0.4314 \\ 0.1044 \\ 0.0328 \end{bmatrix}. \tag{14}$$

In each of the secondary level indicators, (1) The proportion of capital flow is more important than the asset liability ratio, and obviously more important than the interest cover ratio; asset liability ratio is more important than interest cover. (2) The return on net assets is slightly more important than the operating profit margin, more important than the surplus cash guarantee ratio, and significantly more important than the cost utilization ratio; the operating profit margin is more important than the surplus cash guarantee ratio and obviously more important than the cost utilization ratio; surplus cash cover ratio is more important than cost utilization ratio; (3) The R & D investment ratio is slightly more important than the employee salary growth rate, more important than the performance reward ratio, and significantly more important than the equity reward ratio; the wage growth rate of employees is more important than the performance reward ratio and significantly more important than the equity reward ratio; performance reward ratio is more important than equity reward ratio; (4) The proportion of financial errors is important to the quality of financial personnel and extremely important to the proportion of financial personnel in the company members.

The secondary index weight of each primary index can be obtained by using a similar method. They are $U_1 = [0.7306\ \ 0.1884\ \ 0.081]^T$, $U_2 = [0.5338\ \ 0.3101\ \ 0.1135\ \ 0.0426]^T$, $U_3 = [0.5338\ \ 0.3101\ \ 0.1135\ \ 0.0426]^T$, and $U_4 = [0.7219,\ 0.2271,\ 0.051]^T$, respectively. The weight of each secondary index can be calculated by the following formula:

$$k = a_i \times U_j, \quad i = 1, 2 \ldots, 4\ j = 1, 2, \ldots 4, \tag{15}$$

where a_i is the weight of the ith element in vector a and Uj is the index weight vector of the secondary indicators in the jth primary.

3.3. An EFRM Optimization. Based on the above introduction to a company's financial management system to select the solvency, profitability, profit distribution, and FRM ability of four primary financial risk indexes to evaluate a company's FRM ability, Figure 4 shows the four primary financial risk index scores of the enterprise in recent 5 years.

Figure 4 shows that the EFR of the enterprise have been constantly changing in the past five years. (1) affected by the outbreak of the new champions league in 2020, there is the sharp decline in profitability and solvency of enterprises, EFR; (2) In 2018, the company had the lowest financial management ability, which had a negative impact on the normal operation of the company. Later, the company increased its attention and investment in financial management ability and (3) in the process of enterprise operation, all kinds of risks are intertwined and resist risk together to all enterprises.

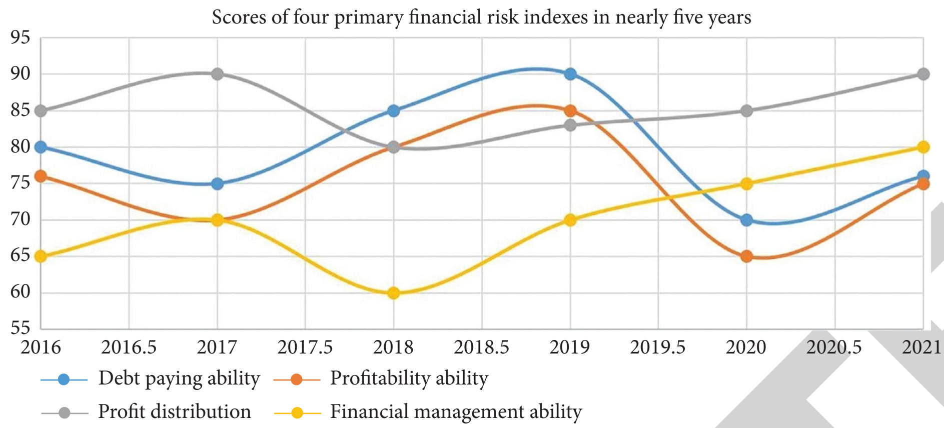


FIGURE 4: Scores of four primary financial risk indexes in nearly five years.

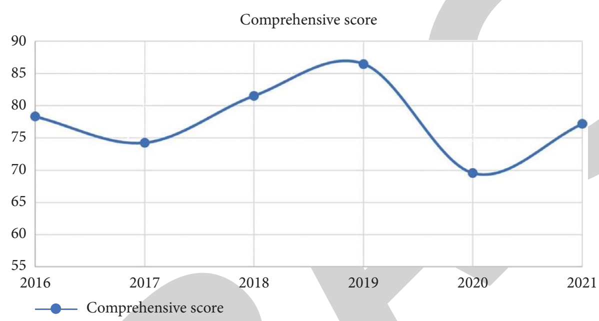


FIGURE 5: Nearly five years against financial risk score.

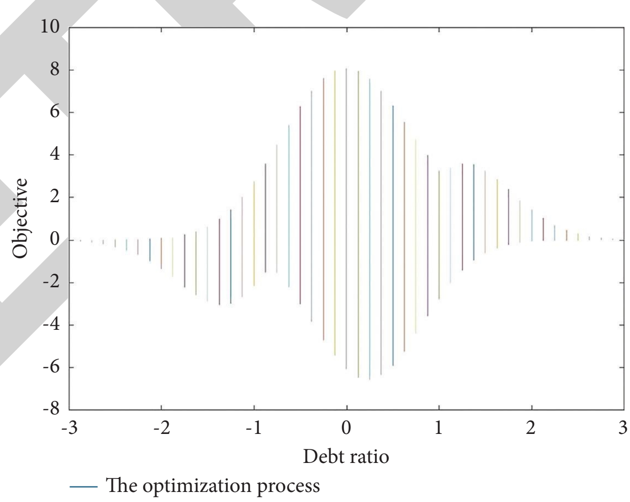


FIGURE 6: Relationship between debt ratio and the multiobjective optimization function.

Based on the above, the enterprise financial risk model and the enterprise financial risk assessment model are established, and the scores of the enterprise's ability to resist financial risks in recent 5 years are obtained according to formula (14), which is shown in Figure 5.

Figure 5 shows that (1) the enterprise to resist the ability of the financial risk is the outcome of combined action of various indicators of risk; (2) the enterprise’s profit ability and debt paying ability to resist the ability of the financial risk of the enterprise had the greatest influence, thus affected by the outbreak of the new champions league in 2020, nearly 5 to the financial risk of the largest companies face; (3) enterprise financial management ability is the ability to resist financial risks to the enterprise minimal impact factor and is

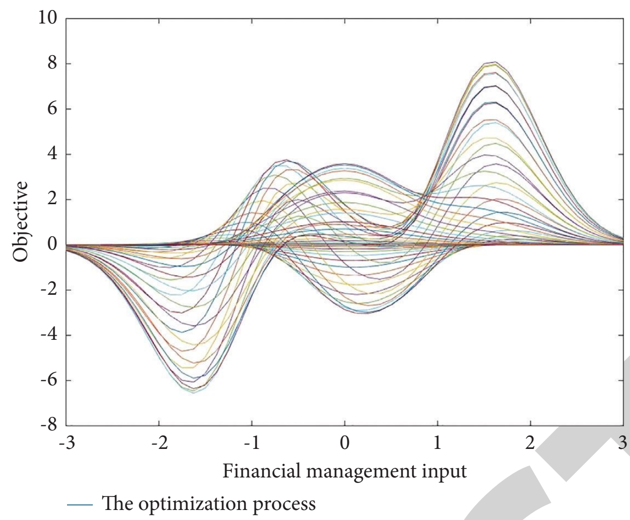


Figure 7: Relationship between financial management input and multiobjective optimization function.

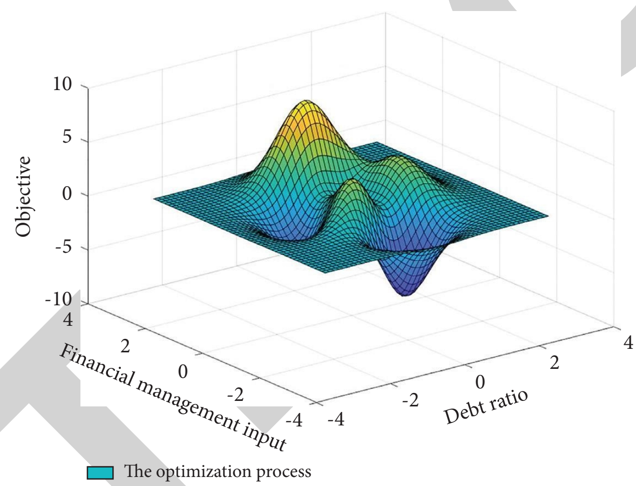


Figure 8: Relationship between bivariate joint and multiobjective function.

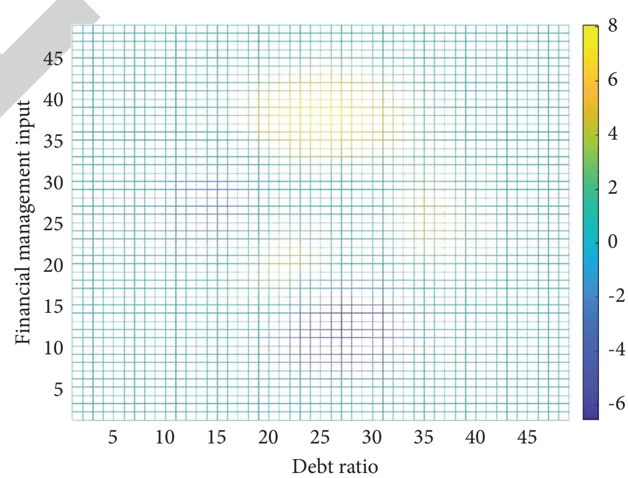


Figure 9: Multiobjective optimization results.

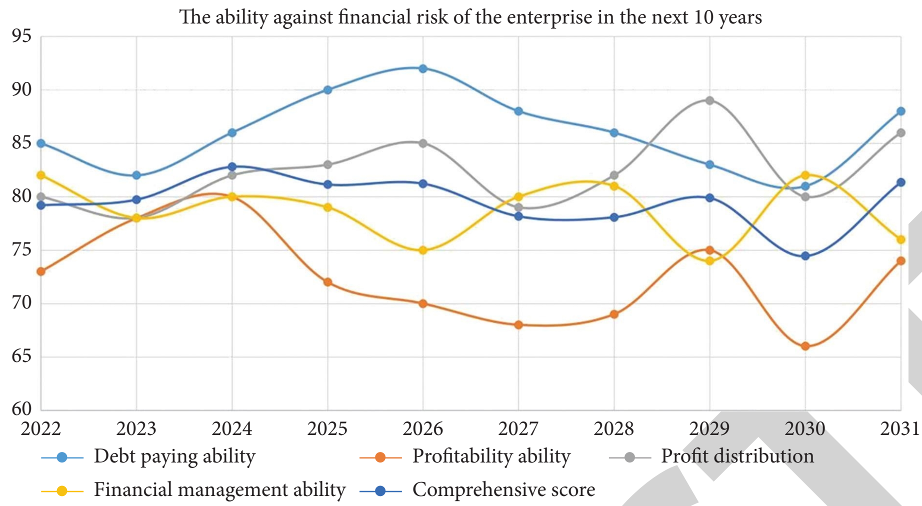


Figure 10: The ability against financial risk of the enterprise in the next 10 years.

often ignored by business leaders, but it is long to the enterprise that the whole operation plays an important role, and attention needs to be improved.

Figures 6–10 show the optimization process for multiobjective optimization of enterprise financial management based on the optimization algorithm shown in Figure 3.

Figure 6 shows the EFRM in the process of multiobjective optimization, debt ratio, and the relationship between the multiobjective optimization function. It shows that when the debt ratio exceeds 1, the multiobjective function value of the enterprise's financial risk gets a smaller value. At this time, the enterprise's ability to resist financial risk is weak. Therefore, such situations should be avoided in the process of enterprise operation.

Figure 7 shows the EFRM in the process of multiobjective optimization and the financial management of the relationship between the input and multiobjective optimization function.

Figure 8 shows the EFRM in the process of multiobjective optimization, selection of debt ratio and financial management into two variables, the two variables, and the relationship between the multiobjective optimization function.

Figure 9 shows that the enterprise’s FRM multiobjective optimization is to find a local optimum, increase investment in certain aspects of the part that can make the target index improved, and some deterioration; this requires us to have the vision of EFRM. According to the conclusion in Figure 9, a local optimization node optimizes the enterprise financial management; the optimized enterprise ability to resist financial risk score to predict the future 10 years is shown in Figure 10.

Figure 10 shows the optimized process for the EFRM, to predict the ability of the enterprise to resist financial risks in the next 10 years.

4. Conclusions

Research materials about EFRM are summarized; the development research history, status, and the current research hot spots about the EFRM are combed in this paper. According to the characteristics of EFRM, the multiobjective optimization method is selected to solve the difficult and hot issue of EFRM. A multiobjective optimization model of EFRM is studied and established, and the intelligent algorithm is used to solve the multiobjective optimization problem. Finally, a company is taken as a research object, and the theory of the research achievements of the EFRM is adopted to optimize the multiobjective optimization model. In order to improve competitive ability of the enterprise market to against financial risk, it provided several suggestions. And if enterprises adopt these suggestions, the EFR of the company will reduce in a certain level in the next 10 years. Therefore, the researching results show that these suggestions to the enterprise play a very important and positive role to improve ability to resist financial risk. The theory and method studied in this paper can provide a theoretical guidance to improve the ability of companies to resist financial risks.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Jiangsu Province Philosophy and Social Sciences in university 2020, “the Research on the Effect of the Financial and Taxation Policy Support under the Pandemic in Jiangsu Province” (project number: 2020SJA2332). The paper is the periodical achievement of “the Accounting Major Labor Education Innovation Team” (the Team Leader: Weiwei Jiang) which was Funded by the Teacher Innovation Team of Suzhou Higher Vocational Education; Zhejiang Province Natural Science Foundation

2020, "The Mechanism Study of the Effect on the Suffering Consciousness Manager to the Cash Holdings" (project number: LY21G020006).

References

[1] B. Josenildo, S. Renato, and B. JoséArnaldo, "Perspectives and relationships in Supply Chain Simulation: a systematic literature review," *Simulation Modelling Practice and Theory*, vol. 62, no. 1, pp. 166–191, 2016.

[2] M. Amini, T. Wakolbinger, M. Racer, and M. G. Nejad, "Alternative supply chain production–sales policies for new product diffusion: an agent-based modeling and simulation approach," *European Journal of Operational Research*, vol. 216, no. 2, pp. 301–311, 2012.

[3] J. He, Y. Huang, and D. Chang, "Simulation-based heuristic method for container supply chain network optimization," *Advanced Engineering Informatics*, vol. 29, no. 3, pp. 339–354, 2015.

[4] C. Y. Li, J. Li, H. L. Chen, M. Jin, and H. Ren, "Enhanced Harris hawks optimization with multi-strategy for global optimization tasks," *Expert Systems with Applications*, vol. 185, pp. 1–30, Article ID 115499, 2021.

[5] B. Cao, J. Zhao, Y. Gu, S. Fan, and P. Yang, "Security-aware industrial wireless sensor network deployment optimization," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5309–5316, 2020.

[6] S. Roy and A. Gupta, "Safety investment optimization in process industry: a risk-based approach," *Journal of Loss Prevention in the Process Industries*, vol. 63, pp. 1–10, Article ID 104022, 2020.

[7] E. B. Abrahamsen and F. Asche, "The insurance market's influence on investments in safety measures," *Safety Science*, vol. 48, no. 10, pp. 1279–1285, 2010.

[8] H. Akbulut and E. Secilmiş, "Estimation of a social discount rate for Turkey," *Socio-Economic Planning Sciences*, vol. 67, no. 1, pp. 78–85, 2019.

[9] D. A. Mellichamp, "Internal rate of return: good and bad features, and a new way of interpreting the historic measure," *Computers & Chemical Engineering*, vol. 106, no. 1, pp. 396–406, 2017.

[10] D. A. Mellichamp, "Exo-parametric ("inside-out") model of discounted cash flow calculations using NPV%: m," *Computers & Chemical Engineering*, vol. 119, no. 1, pp. 309–314, 2018.

[11] D. A. Mellichamp, "Profitability, risk, and investment in conceptual plant design: optimizing key financial parameters rigorously using NPV," *Computers & Chemical Engineering*, vol. 128, no. 1, pp. 450–467, 2019.

[12] J. Lu and X. Chen, "Risk model of financial supply chain of internet of things enterprises: a research based on convolutional neural network," *Computer Communications*, vol. 183, no. 1, pp. 96–106, 2022.

[13] B. Denkena, M. A. Dittrich, and S. Stamm, "Dynamic bid pricing for an optimized resource utilization in small and medium sized enterprises," *Procedia CIRP*, vol. 67, no. 1, pp. 516–521, 2018.

[14] M. Matsumoto and S. Komatsu, "Demand forecasting for production planning in remanufacturing," *International Journal of Advanced Manufacturing Technology*, vol. 79, no. 1-4, pp. 161–175, 2015.

[15] M. Zhang, Y. Tang, L. Liu, and D. Zhou, "Optimal investment portfolio strategies for power enterprises under multi-policy scenarios of renewable energy," *Renewable and Sustainable Energy Reviews*, vol. 154, pp. 1–12, Article ID 111879, 2022.

[16] P. Rodrigo, W. David, and N. Matías, "Portfolio applications in electricity markets review: private investor and manager perspective trends," *Renewable and Sustainable Energy Reviews*, vol. 81, no. 01, pp. 192–204, 2018.

[17] F. Cucchiella, M. Gastaldi, and M. Trosini, "Investments and cleaner energy production: a portfolio analysis in the Italian electricity market," *Journal of Cleaner Production*, vol. 142, no. 1, pp. 121–132, 2017.

[18] J. B. Oliveira, M. Jin, R. S. Lima, J. Kobza, and J. Montevechi, "The role of simulation and optimization methods in supply chain risk management: performance and review standpoints," *Simulation Modelling Practice and Theory*, vol. 92, no. 1, pp. 17–44, 2019.

[19] G. Soni and R. Kodali, "A decision framework for assessment of risk associated with global supply chain," *Journal of Modelling in Management*, vol. 8, no. 1, pp. 25–53, 2013.

[20] I. Heckmann, T. Comes, and S. Nickel, "A critical review on supply chain risk – definition, measure and modeling," *Omega*, vol. 52, no. 1, pp. 119–132, 2015.

[21] T. Aven and Y. Hiriart, "The use of a basic safety investment model in a practical risk management context," *Reliability Engineering & System Safety*, vol. 96, no. 11, pp. 1421–1425, 2011.

[22] L. Talarico and G. Reniers, "Risk-informed decision making of safety investments by using the disproportion factor," *Process Safety and Environmental Protection*, vol. 100, no. 1, pp. 117–130, 2016.

[23] G. Reniers and T. Brijs, "Major accident management in the process industry: an expert tool called CESMA for intelligent allocation of prevention investments," *Process Safety and Environmental Protection*, vol. 92, no. 6, pp. 779–788, 2014.

[24] K. Govindan and M. Fattahi, "Investigating risk and robustness measures for supply chain network design under demand uncertainty: a case study of glass supply chain," *International Journal of Production Economics*, vol. 183, no. 1, pp. 680–699, 2017.

[25] W. Ye and F. You, "A computationally efficient simulation-based optimization method with region-wise surrogate modeling for stochastic inventory management of supply chains with general network structures," *Computers & Chemical Engineering*, vol. 87, no. 1, pp. 164–179, 2016.

[26] F. Zhang, D. Johnson, M. Johnson, D. Watkins, R. Froese, and J. Wang, "Decision support system integrating GIS with simulation and optimisation for a biofuel supply chain," *Renewable Energy*, vol. 85, no. 1, pp. 740–748, 2016.

[27] H. Rodgers, M. Srivastava, K. S. Pawar, and J. Shah, "Supply chain risk management in India – practical insights," *International Journal of Logistics Research and Applications*, vol. 19, no. 4, pp. 278–299, 2016.

[28] O. Khan and B. Burnes, "Risk and supply chain management: creating a research agenda," *International Journal of Logistics Management*, vol. 18, no. 2, pp. 197–216, 2007.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9874054, 1 page
https://doi.org/10.1155/2023/9874054

Retraction

Retracted: The Relevance of Microblog Sentiment Analysis on College Students' Growth Development and Management

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Wang, "The Relevance of Microblog Sentiment Analysis on College Students' Growth Development and Management," *Security and Communication Networks*, vol. 2022, Article ID 1797080, 8 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1797080, 8 pages
https://doi.org/10.1155/2022/1797080

WILEY | Hindawi

Research Article

The Relevance of Microblog Sentiment Analysis on College Students' Growth Development and Management

Lan Wang

School of Marxism, Weinan Normal University, Weinan 714099, Shaanxi, China

Correspondence should be addressed to Lan Wang; wanglan@wnu.edu.cn

Received 28 March 2022; Accepted 25 May 2022; Published 7 June 2022

Academic Editor: Chin-Ling Chen

Most of the existing studies focus on the analysis of college student users of blogs, virtual communities, and other online application platforms, while the research on the motivation of micro-blogging college student users focuses on the research of usage motivation. In this paper, we established a multilevel psychological early warning model based on personality, mood, and emotion space and mapped personality, mood, and emotion to accurately simulate the law of human emotion change for the relevance of micro-blog sentiment analysis to college students' growth development and management. The experiment shows that the method effectively realizes the analysis and description of micro-blog emotion; finally, the designed micro-blog emotion early warning system realizes the visual analysis and timely warning of the psychological condition of the observed subjects, which verifies the effectiveness of the research method.

1. Introduction

With the development of Internet, social networks represented by microblogs have become an important platform for obtaining personal life trajectories and emotional information. The role of microblogs in psychological early warning is getting more and more attention from psychologists in colleges and universities [1]. With the help of computer to analyze the content of microblogs, especially for the key students' microblogs, it can help grasp the recent psychological condition and emotional demands of the observed students and help the college psychologists provide effective interventions for the predicted results to avoid the occurrence of students' psychological crisis [2].

Since the birth of network service and micro-blogging service in October 2006, Twitter has been ranked among the top 15 websites in Europe and the United States in terms of average daily visits and is currently one of the ten most visited websites on the global Internet. As a typical web application in the web 2.0 environment, micro-blogging has gathered a large number of college students with its low-cost content production method and viral spread based on trust chain, which has realized fragmentation, mobilization, real-time, and socialization and has rapidly grown into a new low-cost and most popular media [3]. Therefore, this paper attempts to study the motivation of college students' user-generated content by taking micro-blogging as an example, in order to answer the reasons why college students are so enthusiastic about micro-blogging, to understand the relationship between the motivation and participation behavior of online college students' user-generated content using micro-blogging, to provide some theoretical basis for revealing the inner driving law and motivation of college students' user-generated content behavior on micro-blogging, to obtain hot topics and issues on the Internet, and to monitor the Internet public opinion. It is also a useful reference to help online social tool service providers develop products and services centered on college students and guide the healthy development of microblogs [4].

Research on micro-blogging at home and abroad has only just started, mostly staying in qualitative research, while quantitative research is very rare. Human psychological changes are expressed through emotions, and emotions are also an ever-changing process that is influenced by a variety of factors, such as external stimuli and mood swings [5]. Therefore, building a suitable computer model to simulate human emotion changes is a very challenging topic. The researcher's understanding of the motivation of micro-blog users' content generation is still at the stage of qualitative prediction, lacking quantitative empirical analysis. In an environment where the Chinese micro-blog user base is mature and stable and can provide a large sample size, it is undoubtedly appropriate and important to conduct a quantitative empirical study on the motivation of micro-blog users' content generation [6].

Emotion recognition research in Chinese is relatively late compared to English emotion research, and there is no English general knowledge database such as Concept Net available, and the number of Chinese test corpora for machine learning is too small [7]. For emotion recognition research on short texts such as microblogs (less than 140 characters), it is difficult to extract feature values using machine learning and Bayesian classification algorithms, and the recognition effect is poor, so a complete emotion lexicon with professional domain characteristics needs to be built and applied to the text emotion recognition process. The process of Chinese emotion recognition also involves research areas such as word division, lexical annotation, and syntactic analysis [8].

In this paper, on the basis of the existing research results, the following work is carried out.

(1) Starting from the theories related to emotional psychology, we take emotion, personality, and mood as important factors affecting psychological changes, establish emotion space, mood space, and personality space, reoptimize the relationship between the three for research needs, and propose a multilayer psychological early warning model based on personality, mood, and emotion space. Through this model, the psychological change pattern of human is simulated to provide reliable prediction.

(2) In terms of micro-blogging emotion recognition, a micro-blogging mental emotion dictionary (abbreviated as MPED dictionary) is constructed as the basis of micro-blogging emotion recognition based on the existing emotion dictionary combined with the characteristics of the psychological counseling field. In addition, the emotion meta-theory is proposed, which simplifies the process of emotion recognition of text into the extraction and statistical process of emotion meta. The processes such as word separation and lexical annotation of Chinese were done with the help of ICTCLAS, a word separation software developed by the Chinese Academy of Sciences.

(3) A prototype system was established to verify the effectiveness of the research method in this paper, which can help university psychologists reduce the labor intensity of manual recognition and provide timely early warning of psychological crisis.

The experiment shows that the method effectively realizes the analysis and description of micro-blog emotion; finally, the designed micro-blog emotion early warning system realizes the visual analysis and timely warning of the psychological condition of the observed subjects, which verifies the effectiveness of the research method.

The rest of this paper is as follows: Section 2 is related work, and we summarize the latest research. Section 3 describes our proposed psychological early warning model, and we have carried out mathematical modeling of personality space. Section 4 is the emotion meta model. Section 5 is the experimental content. Section 6 is the conclusion.

2. Related Work

There have been many foreign scholars who have conducted relatively in-depth theoretical and case studies on the motivation of college users to use social networking sites (SNS, Social Network Sites) such as blogs and microblogs. [9] has established a theoretical framework for their research or investigated the behavioral habits of a certain group of people through empirical studies. [10] showed that college users' self-confidence or self-efficacy, need to belong, self-construal, and collective self-esteem have a positive impact on college users' use of SNS sites or online tools. The positive effects of SNSs and online tools on college students' use of SNSs and online tools were suggested by [11, 12] which proposed that the factors of risk, privacy, and security perceived by college users in social networks have an impact on college users' participation behavior. [13] analyzed the content and form of blogs used by college users and found that the main motives of blogging college users use blogs are in the areas of emotional confession and expression. [14–16] concluded that the same topics and themes, a strong sense of personal belonging and community, the same life experience and social experience, and the identification of self-values with the values of others are the main factors influencing the motivation of college users. The content analysis of the blogs in the near-Polish language family outlined that self-expression, interactivity, entertainment to spend time, information gathering, and improving professional competence are the motivations of the college users of blogs for writing. In the study of blogging by university users, the main motivations were sharing experiences, recording daily life and expressing opinions, and using methods such as in-depth interviews and text analysis. By using the technology acceptance model theory, it was concluded that the influence of intrinsic factors on college users' use of microblogs is greater than that of extrinsic factors on college users' use of microblogs. The HITS algorithm was used to effectively categorize college users, and then the study concluded that college users use microblogs for the purpose of sharing their daily speech and collecting and sharing valuable information. [17–19] conducted an empirical study on the motivation of college students to write blogs, and the four factors proposed were gaining fame and fortune, emotional expression, information sharing, and interpersonal communication. [20] analyzed and integrated four factors: emotional motivation, informational motivation, social motivation, and recorded expression motivation in measuring the motivation indicators of college users using microblogs.

3. Psychological Early Warning Model

3.1. Emotional Space. Currently dominating the psychological and engineering communities are six emotion classifications, that is, happy, angry, disgusted, fearful, sad, and surprised. In the field of psychological counseling, whether a visitor has negative (negative) emotion and the duration of negative emotion is a very important basis for assessing the psychological condition of a visitor, the research in this paper focuses on the identification of negative emotion. At present, the research on Chinese text emotion recognition is still in its initial stage, and most of the studies focus on positive and negative dichotomy of emotion. Combined with the practical needs of this paper, a two-dimensional emotion space is constructed, and its affective variable is $E = [e_{\text{pos}}, e_{\text{neg}}], \forall_{e[.]} \in [0.1]$ (e_{pos} for positive emotions and e_{neg} for negative emotions). All 6 emotions mentioned in the literature [8] can also be classified into these 2 intervals, namely, negative emotions-anger, disgust, fear, sadness, surprise, and positive emotions-High.

3.2. Mood Space. The PAD three-dimensional mood space model proposed by [9] was used, where P denotes Pleasure, A denotes Arousal, and D denotes Dominance, and these three-dimensional traits are independent of each other and constitute the three-dimensional mood space. Define the three-dimensional mood space variable $M = [m_P, m_A, m_D]^T, -1 \le m_P, m_A, m_D \le 1,$ where $M = [0, 0, 0]^T$ corresponds to the mood of the calm state.

3.3. Personality Space. The Five Factor Model (FFM) is a very widely used personality model with five factors representing openness, responsibility, extraversion, agreeableness, and neuroticism, respectively. In this paper, the five-factor model is used as a personality model, and personality is represented as a five-dimensional vector $P = [p_1, p_2, p_3, p_4, p_5]^T$, where, $p_i \in [0, 1]$, $i = 1, 2, 3, 4, 5$.

3.4. Personality, Mood, and Emotion Interrelationship. Since mood space is a three-dimensional space, each dimension takes positive and negative values, respectively, while personality space is a five-dimensional space; for this reason, the personality and mood space conversion relationship is established as $M = KP$. where M is the mood intensity; P is the personality; K is the personality and mood conversion matrix; that is,

$$K = \begin{bmatrix} 0 & 0 & 0.21 & 0.59 & 0.19 \\ 0.15 & 0 & 0 & 0.30 & -0.57 \\ 0.25 & 0.17 & 0.60 & -0.32 & 0 \end{bmatrix}. \tag{1}$$

Changes in emotion are not only related to external stimulus signals and personality, but also influenced by the current mood. Referring to the three-dimensional mood PAD space and emotion quantification relationship table proposed in literature [11], the correspondence is listed in Table 1, combined with the research characteristics of this paper.

Table 1: Correspondence between PAD space and emotion space.

Emotion	Pleasure	Arousal	Dominance	Mood subspace
Positive emotion	0.40	0.20	0.15	+P + A + D
Negative emotion	−0.40	−0.20	−0.50	−P − A − D

According to Table 1, the mapping matrix of mood and emotion is defined as

$$L = \left[e_{\text{pos}}, e_{\text{neg}}\right] = \begin{bmatrix} 0.40 & -0.40 \\ 0.20 & -0.20 \\ 0.15 & -0.50 \end{bmatrix}. \tag{2}$$

Define the mapping relationship between the affective and mood variables as

$$\begin{gathered} \mathbf{E} = f(\mathbf{M}, \mathbf{L}) = \frac{\mathbf{D}}{d_{\text{pos}} + d_{\text{neg}}}, \\ \mathbf{D} = \left[d_{\text{pos}}, d_{\text{neg}}\right], \quad d_i = \left[(\mathbf{M} - e_i)^T (\mathbf{M} - e_i)\right]^{1/2}, \end{gathered} \tag{3}$$

where $i = 1, 2,$ corresponding to positive emotion and negative emotion, respectively.

3.5. Early Warning Model Establishment. By mining the emotion information in the micro-blogging text of the observed subjects and converting the emotion information into emotion evoking variables, which act on the mood space and emotion space, respectively, the possible changes of emotion and mood under the influence of personality factors are predicted, and the system provides crisis warning when the emotion variables exceed the warning threshold.

Mood state changes are influenced by the evoked variables and personality, so the mood update equation can be expressed as

$$M_t = M_{t-1} + KP_t + \varphi(A_t, P_t) + \xi(P_t, M_{t-1}), \tag{4}$$

where M_{t-1} is the mood state at the moment $t-1$; KP_t is the component of the influence of personality P_t on mood state; $\varphi(A_t, P_t)$ is the component of the influence of affective evoked variables on mood; and $\xi(P_t, M_{t-1})$ is the mood decay component, which plays an inhibitory role in mood change. In this paper, we assume that the evoked variables are linearly related to the mood state variables with the following equation:

$$\varphi(A_t, P_t) = \omega_m L A_t, \tag{5}$$

where ω_m is the coefficient of influence of external stimuli on mood fluctuations, determined by personality, and A_t is the external evoked variable.

After the external stimulus disappears, the mood gradually decays over time and eventually returns to a calm state M_0. The mood decay equation is described as follows:

$$\xi(P_t, M_{t-1}) = -\frac{\alpha(M_{t-1} - M_0)}{1+\alpha}, \quad (6)$$

where α is the mood decay factor, the size of which depends on personality.

Changes in affective state are influenced by the predisposing variables, personality, and mood factors, and the following affective update equation is established.

$$E_t = E_{t-1} + f(M_t, L) + \varphi(A_t, P_t) + \phi(P_t, E_{t-1}), \quad (7)$$

where $\varphi(A_t, P_t)$ is the component of the effect of emotion-inducing variables on affective states, $\varphi(A_t, P_t) = \omega_e A_t, \omega_e$ is the emotion-inducing factor, which is related to personality, and $\phi(P_t, E_{t-1})$ is the decay component of emotion, which is defined after the mood decay component as follows:

$$\phi(P_t, E_{t-1}) = -\frac{\beta(E_{t-1} - E_0)}{1+\beta}, \quad (8)$$

where E_0 is the emotional calm state; β is the emotional attenuation factor, which is related to personality.

4. Emotion Meta-Model

Positive and negative emotion words and sensitive words in the field of psychological counseling are selected to build a special micro-blog mental emotion dictionary (MPED dictionary for short). In the rule-based approach based on emotion dictionary, emotion words are the most important consideration in micro-blog emotion recognition. Relying on individual emotion words alone, ignoring the relationship between words in context and the negation words and degree adverbs that have important influence on emotion tendency and intensity, it is easy to produce great ambiguity, so the concept of emotion element is proposed. Emotion meta is the most basic component of micro-blogging emotion and is the basic unit of emotion judgment for phrases, sentences, and sections. The proposed concept of sentiment meta-sense simplifies the process of judging the sentiment tendency of phrases, sentences, and sections and makes the whole reasoning simpler, clearer, and more explicit, which is defined as follows:

$$C_e = (E_m, N, Q, L_a). \quad (9)$$

The set of 1 C_e is the finite set of sentiment elements, $C_e = \{C_0, C_1, C_2, \ldots, C_y\}, y$ is the number of sentiment elements, E_m is the set of sentiment words, $E_m = \{E_{m0}, E_{m1}, E_{m2}, E_{m3}, E_{m4}\}$ (5 types of sentiment words: positive, negative, neutral, domain positive, and domain negative, respectively), N is the set of negation words, $N = \{N_0, N_1\}$ ("with" and "without" negation words, respectively), Q is the set of degree adverbs, $Q = \{Q_0, Q_1, Q_2, Q_3\}$ (4 types of degree adverbs: "slightly," "more," "more," and "most," respectively), L_a is the position of degree adverbs, and $L_a = \{L_{a0}, L_{a1}\}$ negation words and degree adverbs are the adjacent words of the sentiment words.

To verify the validity of the research in this paper, a micro-blog emotion recognition system is designed to identify the emotional state of the observed object through the micro-blog text [21, 22]. The system algorithm is built on the basis of the text analysis algorithm of MPED dictionary, and the basic steps are as follows.

(1) Extract the original micro-blog text from the web page and store it in the local database.
(2) Preprocess the original text, including word separation and lexical annotation.
(3) Construct MPED dictionaries (consisting of How Net sentiment dictionaries, student praise and derogatory dictionaries, etc., and dictionaries of sensitive words in the field of psychological counseling).
(4) Extract sentiment elements and constructing a finite-state automaton to analyze the sentiment elements' tendencies.
(5) Sentence analysis is performed to determine the overall sentiment tendency of the sentences, and the overall sentiment output of the microblogs is further derived by accumulation.
(6) The overall sentiment output of tweets is input to the early warning model as a predisposing variable.
(7) The current possible sentiment state of the observed object is calculated and displayed through the trend graph output. If the warning threshold value is exceeded, a crisis alert is stimulated.

The system structure is shown in Figure 1.

5. Experiment and Analysis

The hardware used in this experiment is a Lenovo Y500 notebook, the CPU model is i5-3210m, and the memory is 8 GB. The software platform is Python 3.5.3, which mainly uses three libraries: pandas, numpy, and sklearn. The hyperparameter is set to batch size of 128, and the learning rate is 0.001. Adam optimizer is used.

The correct classification rate is used to evaluate the effectiveness of classification. Let the number of positive tweets in the test dataset be N_P, the number of negative tweets be N_n, the number of warning tweets be N_w, the number of positive tweets in the system classification be Y_p, the number of negative tweets be Y_n, and the number of warning tweets be Y_w, and then the accuracy rate is calculated as follows.

The experiment takes 2 methods for simulation testing.

5.1. Method 1. Under the condition that the observation subject's personality and current mood state were not available, the ideal state was selected, and a single experiment was conducted on the micro-blog test texts to count the system prediction accuracy [23, 24].

The experimental data were obtained from an automatic micro-blog text collection program developed by the author, and 2,000 texts were downloaded from Sina Weibo as the test corpus, of which 1,000 were positive, and 1,000 were negative, and 226 were identified manually to reach the warning level. The average accuracy rate of positive tweets is

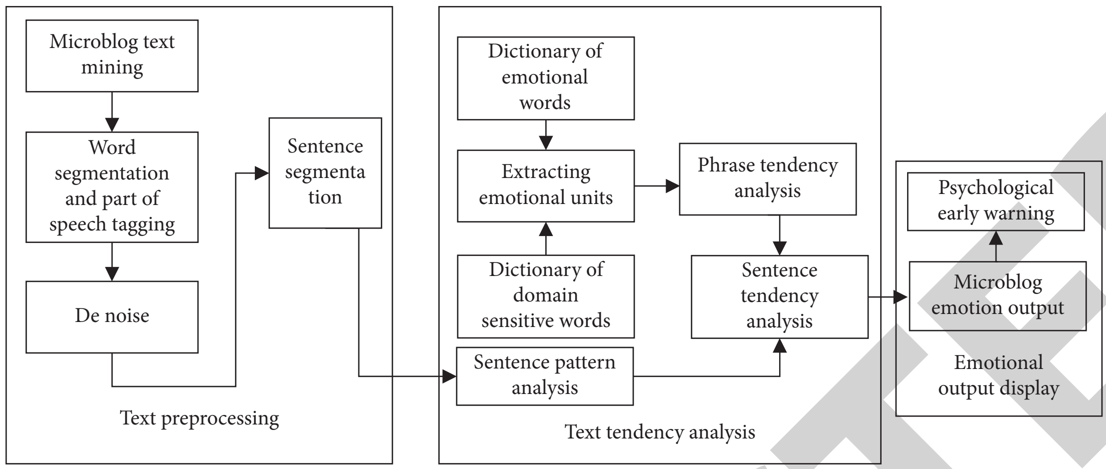


Figure 1: System structure.

Table 2: Comparative experimental results.

Type	Manual identification	MPED dictionary	
	Number	Number	Accuracy (%)
Positive micro-blog	1000	746	74.6
Negative micro-blog	1000	738	73.8
Early warning micro-blog	226	723	72.3

81%, while the average accuracy rate of negative tweets is 78%.

It can be seen that, based on manual identification, the recognition rate of micro-blog sentiment is above 78%, and the accuracy rate of early warning reaches 85%. Most of the reasons for the errors are focused on the limitations of the MPED dictionary and the arbitrary nature of online language; that is, the errors that occur during the input of micro-blog sentiment are brought into the early warning model.

5.2. Method 2. Multiple experiments were conducted on the microblogs of 100 school students with known personality and other factors; that is, the content of the above students' microblogs was tracked and analyzed for 7 consecutive d. The accuracy of the system's prediction during the consecutive experiments was obtained by a comparative survey on the 7th day. The identification accuracy rates were as follows: 75% for positive tweets and 70% for negative tweets, with an average accuracy of 72% and an early warning accuracy of 78%.

The overall index of prediction accuracy of method 2 has decreased compared with method 1. The reasons for this are that, in addition to the input error, the description of the test subject's personality and mood state is not precise enough, leading to errors in making predictions; the errors will continue to accumulate in the course of continuous experiments, leading to a continuous decrease in the final prediction accuracy. These are the areas that need to be improved in the next step [25, 26].

The above experiments show that the system can better predict the possible emotional changes of the observed objects based on the input of micro-blogging emotions, which greatly reduces the labor intensity of manual identification and provides an effective aid for college psychologists.

The purpose of this experiment is to compare the accuracy of the micro-blog text emotion classification only under ideal conditions (without considering the blogger's personality, mood, and current emotional state) to verify the effectiveness of the improved algorithm based on the emotion dictionary in this paper, and also to lay the foundation for the next simulation test.

Before the classification, the Chinese text was firstly divided into words using the word separation software of the Institute of Computer Science of the Chinese Academy of Sciences, and then the Unigram and bigram of the words were selected as features for the experiments. The experimental results are shown in Table 2.

From the experimental results, the accuracy of the classification is more than 72%, which is a good classification effect. At the same time, there are some problems: due to the limitation of dictionary coverage and the randomness of Internet language, some words and the latest Internet vocabulary cannot be identified in the dictionary, which affects the overall judgment; the grammatical composition of Chinese is extremely complex, the syntactic analysis technology in this paper is not perfect, and the recognition of information needs to be improved; there are some emotional words such as "pride." The construction of a sensitive lexicon in the field of psychological counseling is the first of its kind, and the number of words is small, and the scope of coverage is limited, which needs to be improved in the next step.

The accuracy of the system's prediction during the continuous experiments was obtained by comparing the microblogs of 100 students in our school for 7 days. Students were asked to fill out the Eysenck Personality Inventory and obtain their personality data through statistical analysis. The results of the micro-blogging data collected during the 7-day experiment are shown in Table 3, compared with those of the system and manual recognition.

During the system test for 7 consecutive days, the recognition rate of each index of the system was above 70%, and the recognition rate was highest in the first three days of

Table 3: Simulation results.

Date	Number of microblogs	Manual identification			System prediction					
		Positive	Negative	Early warning	Positive	Accuracy (%)	Negative	Accuracy (%)	Early warning	Accuracy (%)
1	436	280	156	49	222	80.4	126	80.5	38	77.5
2	502	365	141	37	287	80.9	113	79.5	25	78.1
3	388	215	173	55	165	77.6	124	76.5	39	73.3
4	475	173	201	63	202	76.6	160	76.7	46	75.6
5	322	190	132	58	147	75.7	98	74.5	43	73.5
6	309	211	99	46	159	74.8	71	74	34	74.7
7	288	175	116	59	132	71.8	80	71.8	45	72.6
	2694	1609	1018	367	1314	77.5	772	76.6	270	74.50

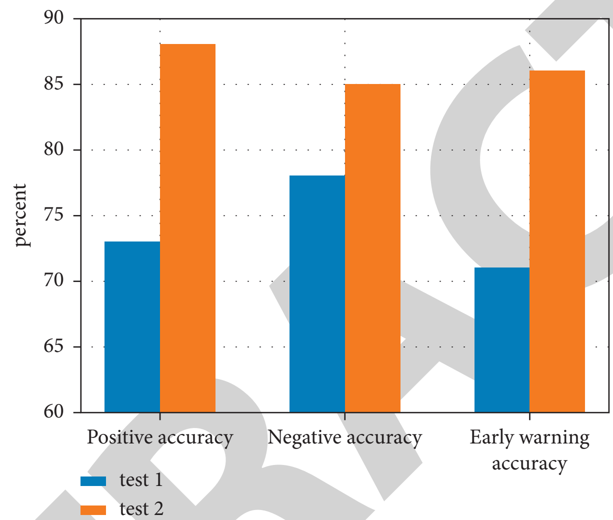


Figure 2: Comparison of experimental results.

operation, and the recognition rate was decreasing in the later period as a whole. The analysis may be the error caused by the imprecise description of the test subject's character and mood condition, as well as the effect accumulated in the operation process, which are the places we need to improve in the next step.

As shown in Figure 2, the accuracy of the predictions of the simulation test has increased compared with that of experiment 1 (ideal state). This indicates that the testers' personality, mood, and emotional condition have a good correction effect on the emotional classification of the micro-blog text, and the psychological early warning model can better simulate the process of human emotion generation and change.

Emotional space trend diagram: after systematic analysis, the emotional space trend of "Pippi Time Machine" in this stage is shown in Figure 3.

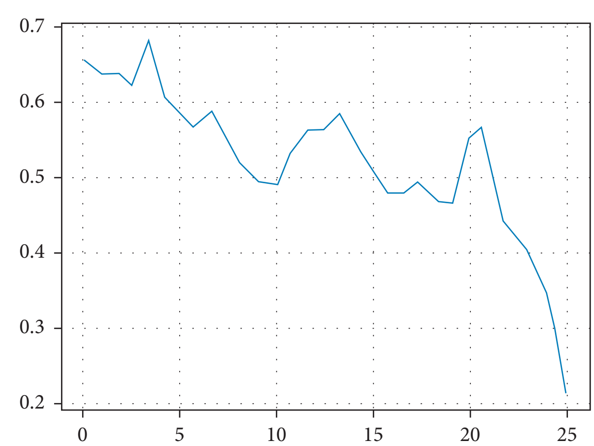


Figure 3: Pippi Time Machine sentiment chart.

The last week of the subject's sentiment trend is in the negative warning level, the negative sentiment value of micro-blogging has reached −0.4 orange warning level since March 12, and the negative sentiment of micro-blogging before "suicide" reached −0.713 red warning level on March 18. If we can take corresponding precautionary measures beforehand, it is possible to avoid such malignant events. Based on the continuity of negative emotions in psychological research, this system can effectively provide the trend of emotions and assist psychologists in making early warning judgments.

6. Conclusion

By analyzing the emotional tendency of micro-blog and starting from the relevant theories of emotional psychology, this paper takes emotion and personality as important factors affecting psychological changes, reoptimizes the relationship between the three, simulates people's psychological change mode through the model, provides reliable prediction, speculates the current emotional state of the observed object, and puts forward a multilevel psychological early warning model based on personality, emotion, and emotional space according to the characteristics of micro-blog. It can accurately simulate the changes of human emotional state; effective meta-theory transforms the text affective analysis into the reasoning and statistical process of affective meta, which makes the analysis more clear; according to the characteristics of micro-blog, combined with the relevant knowledge in the field of psychological counseling, an MPED dictionary is developed. Experiments show that the dictionary-based method can effectively judge the emotional tendency of micro-blog with high accuracy.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: a hybrid system using *n*-gram analysis and dynamic artificial neural network," *Expert Systems with Applications*, vol. 40, no. 16, pp. 6266–6282, 2013.

[2] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for twitter sentiment analysis," *IEEE Access*, vol. 6, pp. 23253–23260, 2018.

[3] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López, and A. R. Montejo-Ráez, "Sentiment analysis in twitter," *Natural Language Engineering*, vol. 20, no. 1, pp. 1–28, 2014.

[4] A. Giachanou and F. Crestani, "Like it or not," *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–41, 2016.

[5] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: a review and comparative analysis of web services," *Information Sciences*, vol. 311, pp. 18–38, 2015.

[6] F. M. A. Abd-AlGalil, S. P. Zambare, and A. M. A. Mashaly, "First record of Chrysomya saffranea (Diptera: Calliphoridae) of forensic importance in India," *Tropical Biomedicine*, vol. 33, no. 1, pp. 102–108, 2016.

[7] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Stream-based active learning for sentiment analysis in the financial domain," *Information Sciences*, vol. 285, pp. 181–203, 2014.

[8] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis-A review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16–32, 2018.

[9] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Systems with Applications*, vol. 77, pp. 236–246, 2017.

[10] B. Ji, Y. Li, D. Cao, C. Li, S. Mumtaz, and D. Wang, "Secrecy performance analysis of UAV assisted relay transmission for cognitive network with energy harvesting," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7404–7415, 2020.

[11] X. Lin, J. Wu, S. Mumtaz, S. Garg, J. Li, and M. Guizani, "Blockchain-based on-demand computing resource trading in IoV-assisted smart city," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1373–1385, 2021.

[12] J. Li, Z. J. Zhou, J. Li et al., "Decentralized on-demand energy supply for blockchain in Internet of things: a microgrids approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1395–1406, 2019.

[13] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Systems with Applications*, vol. 69, pp. 214–224, 2017.

[14] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys*, vol. 50, no. 2, pp. 1–33, 2018.

[15] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Computers in Human Behavior*, vol. 31, pp. 527–541, 2014.

[16] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, 2015.

[17] J. Hunt and D. Eisenberg, "Mental health problems and help-seeking behavior among college students," *Journal of Adolescent Health*, vol. 46, no. 1, pp. 3–10, 2010.

[18] A. Jungherr, "Twitter use in election campaigns: a systematic literature review," *Journal of Information Technology & Politics*, vol. 13, no. 1, pp. 72–91, 2016.

[19] J. L. Hayes, G. Golan, B. Britt, and J. Applequist, "How advertising relevance and consumer-Brand relationship strength limit disclosure effects of native ads on Twitter," *International Journal of Advertising*, vol. 39, no. 1, pp. 131–165, 2020.

[20] B. G. Smith, "Socially distributing public relations: twitter, Haiti, and interactivity in social media," *Public Relations Review*, vol. 36, no. 4, pp. 329–335, 2010.

[21] X. I. E. Tao, C. Zhang, and Y. Xu, "Collaborative parameter update based on average variance reduction of historical gradients," *Journal of Electronics and Information Technology*, vol. 43, no. 4, pp. 956–964, 2021.

[22] Z.-W. Zhang, D. Wu, and C.-J. Zhang, "Study of cellular traffic prediction based on multi-channel sparse LSTM," *Computer Science*, vol. 48, no. 6, pp. 296–300, 2021.

[23] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection," *Information Processing & Management*, vol. 59, no. 2, Article ID 102844, 2022.

[24] W. X. Zhao, J. Jiang, J. He et al., "Topical keyphrase extraction from twitter," in *Proceedings of the 49th annual meeting of the*

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9876421, 1 page
https://doi.org/10.1155/2023/9876421

Retraction

Retracted: Application of CAD in Semifinished Product Packaging Art Design

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Mao and T. Wu, "Application of CAD in Semifinished Product Packaging Art Design," *Security and Communication Networks*, vol. 2022, Article ID 8937896, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8937896, 13 pages
https://doi.org/10.1155/2022/8937896

Research Article

Application of CAD in Semifinished Product Packaging Art Design

Yangyang Mao [1] and Tie Wu [2]

[1] *School of Fashion & Art Design, Xi'an Polytechnic University, Xi'an, Shaanxi 710048, China*
[2] *School of New Media Art, Xi'an Polytechnic University, Xi'an, Shaanxi 710048, China*

Correspondence should be addressed to Yangyang Mao; maoyangyang@xpu.edu.cn

Received 22 February 2022; Revised 18 March 2022; Accepted 1 April 2022; Published 7 June 2022

Academic Editor: Chin-Ling Chen

In order to solve the problem, the packaging CAD software developed in China is mostly the design software of carton and rarely involves the design of cushion liner. It is proposed to use Visual C++ program to develop this software for buffer packaging design, which liberates people from the complicated manual design, significantly shortens the product R&D cycle, and avoids the destructive test of the product. When the overall dimension, quality, brittleness value, cushioning performance parameters of cushioning materials, transportation conditions of packages, and liner parameters are known, the most economical material can be found in the database of cushioning materials through this software, and the liner size structure can be obtained. Based on this data, the structural design of corrugated box is further carried out. For the packaging of products, especially the packaging of precision electronic products, its economic benefits are considerable. Visual C++ 6.0 is used as the software development tool. Using the GUI (graphical device interface) characteristics of various tools and other high-level languages, a simple and friendly user interface is designed to verify the effectiveness of the experiment.

1. Introduction

At present, a lot of work in packaging design can be completed by computer. For example, the structural design of packaging can be completed with AutoCAD, as shown in Figure 1 [1]. The modeling design and surface decoration design of packaging can be completed with Photoshop and 3Dsmax, the stress analysis and buffer packaging design of complex packaging system can be carried out with ANSYS, and several aspects of packaging design can be completed with special packaging CAD. However, most of the packaging design with these methods is only a kind of computer-aided drawing. Whether it is the structural design of packaging or the artistic design of packaging, it is just to show people's design ideas with the help of computers, to help people give up and lose the drawing board. There are still a lot of gaps from the real packaging design system. Moreover, using these current methods for packaging design often requires several kinds of software, the design cycle is long, and the design results cannot be optimized. Therefore, it is necessary to introduce the concept of system and apply artificial intelligence technology to develop a real computer-aided packaging design system [2]. There are many kinds of packaging design, but it should meet the following six requirements. (1) Protection performance: it refers to the ability of packaging to protect various properties of packaging contents, such as moisture resistance , anticorrosion, and shock resistance, and to prevent packaging contents from harming people and the surrounding environment. (2) Promoting art performance: it mainly involves the modeling and decoration of packaging, as well as the function of text advertising, which is often very important for commodity promotion. (3) Service performance: on the one hand, it refers to whether the package is convenient to open, store, and handle and whether it will cause harm to users. On the other hand, it refers to the degree of standardized design and manufacturing adopted. (4) Economy: it mainly refers to whether the comparison between the cost of packaging and the value of content is reasonable. (5) Realizability: it refers to whether the packaging materials are easy to obtain, whether the packaging structure is feasible, whether the whole packaging is suitable for mechanized and automatic

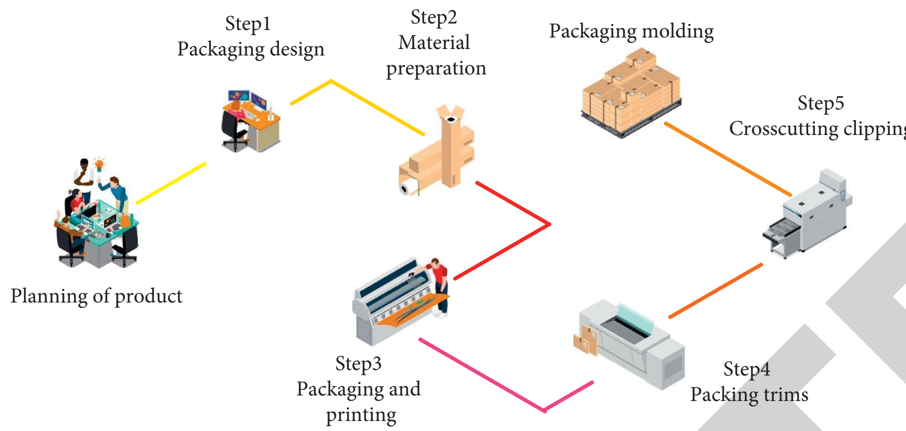


Figure 1: Flowchart of CAD packaging art design.

production, etc. (6) Protection performance: it refers to whether the packaging meets the requirements of environmental protection, such as whether it can be recycled and whether it will cause environmental pollution. In the whole process of packaging design, these requirements should be considered. Computer technology has played an increasingly important role in packaging design, bringing new design methods and concepts to packaging design [2]. The previous computer-aided packaging design is mainly limited to helping designers improve drawing efficiency, which cannot adapt to the development of computer technology. Only by introducing the concept of system and applying artificial intelligence technology to develop a highly intelligent computer-aided packaging design system can computers play a more important role in packaging design [3]. To develop such a system, we need to learn from some mature technologies in other fields, redevelop some special software, and establish the corresponding expert system. It is a huge, complex, and arduous work. However, from the perspective of market, such a computer-aided packaging design system is worth developing.

2. Literature Review

Meng and others found that, in recent years, with the rapid development of China's industrial economy, the improvement of people's quality of life, and the enhancement of environmental protection awareness, paper packaging has become the most potential green packaging form because of its wide source of raw materials, easy molding, degradable recycling, and other advantages [4]. According to the literature, the output value of corrugated paper industry in Sichuan has surpassed that of the packaging industry in Europe since 1999. [5]. According to Yu and Sinigh, the output value of corrugated paper industry in China has also surpassed that of the packaging industry in the world since 2003. Sidelnikov and others found that, in addition to the increasing share of paper packaging in various packaging materials, paper packaging is also developing in the direction of high quality, high strength, lightweight, and versatility, which have also become an important reason for the rapid and stable development of paper packaging industry in the packaging field [6]. With the rapid and stable development of the paper packaging industry, Deja and others have also been widely used in all walks of life in order to meet the needs of multi-Wu chemical industrial design outside China and fully meet the market development needs of small batch, personalization, interest, diversification, and short cycle, it has become a key technology to improve the product R&D capability and enhance competitiveness of various enterprises [7]. Tipsina and others found that, at present, this technology is mainly used in the design of packaging structure, packaging modeling, packaging decoration, packaging machinery, buffer packaging, etc. [8]. Cherevko and others believe that computer-aided technology not only injects new blood into Chinese traditional industry but also makes packaging CAD technology develop rapidly in a scientific and professional direction [9]. As a computer-aided design software widely used in the field of engineering design at home and abroad, Bakhadirov and others believe that it is impossible to meet various industry standards and the usage habits of each user, but AutoCAD has very powerful drawing function and provides rich programming interfaces for the majority of users. This provides very convenient conditions for us to quickly transform it into special software that meets our needs through secondary development in a short cycle [10]. Sahmel and others found that, in the era of rapid changes in science and technology, China's outer packaging industry is taking this opportunity to develop rapidly in the direction of networking, informatization, and intelligence. The existing packaging technology can no longer meet the needs of the development of packaging industry, so it is urgent to research and develop new packaging technology [11]. Golembovskaya and others found that, at present, intelligent packaging has attracted more and more attention because of its intelligent and humanized technical characteristics, extensive social effects, and huge market development potential. From "intelligent control" proposed in 1967 to "expert control" proposed in 1984, the "intelligent packaging" theory in 1992 and the subsequent trend of the Internet of things have the characteristics of one continuous line, which reflects the framework of the system development process of the packaging industry since the information revolution. At the same time, it also reflects that the interdisciplinary and multidisciplinary cross comprehensive application has become the mainstream of the times [12]. Hrudkina and others found that this multidisciplinary and multifield cross integrated application has become the mainstream of information age, and if the packaging industry can grasp the trend of the times and successfully combine the relatively mature high-tech in other disciplines with the existing packaging technology, it will help to improve the overall technical level of the packaging industry [13].

3. Method

The following development principles must be followed in the process of software development.

3.1. Modularization. Module is a relatively independent part of program logic. By means of decomposition, complex problems are divided into thousands of smaller, relatively independent, and easy to solve subproblems in time or scale. Subproblems should have good interface definition and then be solved respectively. For example, functions in C language and classes in C++ language are modules. Modularization helps to abstract and hide information and to represent complex systems [14].

3.2. Abstraction and Information Hiding. Abstraction refers to extracting the most basic features and behaviors of things, ignoring other details irrelevant to the problem. The sharing mechanism of software development process can be improved through hierarchical abstraction and layer-by-layer refinement. Information hiding is to design the module as a "black box," hide the details of data and operation inside the module, and shield the outside world. If users want to access the data in the module, they can only access the outside world through the interface of the module [15]. In this way, the independence of the module can be effectively guaranteed.

3.3. High Cohesion and Low Coupling of Modules. When dividing modules, we should consider concentrating the logically interrelated computer resources into one physical module to ensure loose coupling between modules and strong cohesion within modules. This helps to control the complexity of the solution [16].

3.4. Certainty. The expression of all concepts in the process of software development should be standardized, definite, and unambiguous. This helps people to communicate without misunderstanding and ensure that the whole development work can be carried out in a coordinated and smooth manner [17].

3.5. Uniformity. Consistency is one of the purposes of studying software engineering methods, which is to make software product design follow the guidance of unified and recognized methods and specifications and standardize the development process [18]. It requires that the whole software system (including programs, documents, and data) meet the following consistent characteristics: the concepts, symbols, and terms used are consistent; the internal and external interfaces of the program shall be consistent; the system specification is consistent with the system behavior; consistency of software format; workflow consistency.

3.6. Completeness. Considering the completeness of management and technology is to realize the functions required by the system within the time limit and ensure the software quality. In the process of software development and operation, strict technical review must be carried out to ensure the effectiveness of development results in each development stage.

Cushioning packaging CAD software is based on VC platform. It adopts local cushioning packaging method for products with regular shape (typical cuboid) and products with irregular shape (such as CRT display). At the same time, considering the deviation between product center of gravity and geometric center, it designs cushioning pad and corrugated box and can draw three-dimensional effect drawing of cushioning packaging through OpenGL technology [19]. The software flowchart is as follows (Figure 2).

The cushioning packaging CAD software has the following functions: input the weight, length, width, height, brittleness value, and equivalent drop height of the product to conduct the preliminary design of the buffer pad; input the projection coordinates of the center of gravity of the product, and design the volume, area, and thickness of each buffer pad when the projection of the center of gravity of the product and the geometric center of gravity on the bottom surface does not coincide (i.e., eccentric). According to the needs of users, through the selection of function buttons, the antivibration check, angular drop check, deflection check, creep check, and center of gravity check of buffer pad are carried out, respectively [20]. Select the type of product, refer to the design characteristics in actual production, and adopt OpenGL technology. Draw the three-dimensional effect drawing of buffer packaging scheme, so that users can have an intuitive understanding of the design scheme. Select corrugated type, corrugated box type, and corrugated board code to display the box type structure diagram of the selected corrugated box. The compressive strength can be calculated by the size and the thickness of the corrugated box. Cushioning packaging is a packaging technology that takes certain protective measures to reduce the impact and vibration of the contents of the package from external forces and avoid damage during the circulation of the package. Cushioning packaging is a technical measure to select appropriate cushioning materials and packaging structure to reduce the impact force and vibration force transmitted to the contents to less than the strength of the contents, to ensure the safety and integrity of the contents. The design basis of cushioning packaging is the characteristic parameters of contents, the parameters of circulation environmental conditions, and the parameters of cushioning materials.

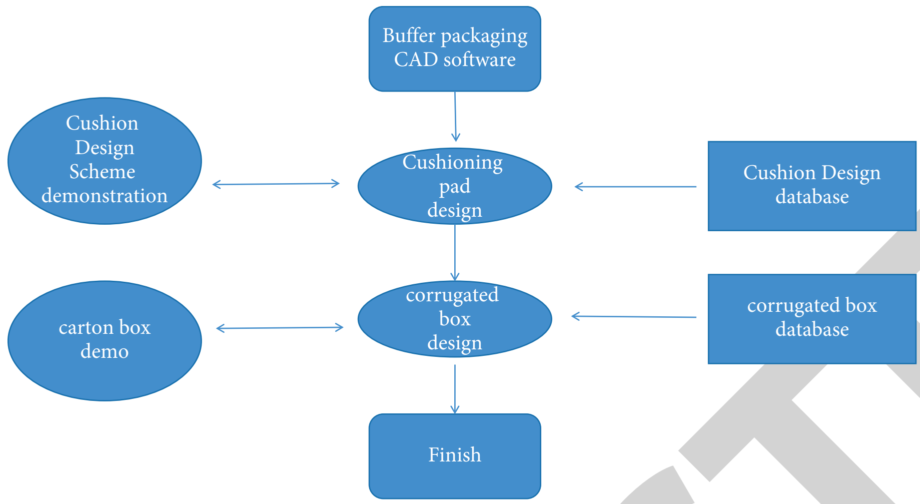


Figure 2: Flowchart of buffer packaging CAD software.

For a specific cushioning package, the characteristic parameters and environmental parameters of the contents are the determined parameter values. Therefore, the main way to require cushioning packaging to achieve the purpose of cushioning and antivibration is to solve it by adjusting the technical measures of cushioning materials and packaging structure. The purpose of cushioning packaging is to protect the performance and shape of the packaged products in case of impact, vibration, and other external forces during transportation and loading and unloading. The design shall meet the following requirements: reducing the impact, vibration, and other external forces transmitted to the products; dispersing the stress acting on the product; protecting the surface and convex part of the product; preventing mutual contact of products; preventing the product from moving in the packaging container; protecting other protective packaging functions.

Embrittlement value is a quantitative index used to express the strength of a product when it is subjected to impact and vibration, also known as the vulnerability of a product. This value indicates the bearing capacity of the product to external force and is the ratio of the maximum acceleration that the package can bear to the gravitational acceleration. It is one of the important parameters of buffer design. The purpose of buffer design is to reduce the external impact force below g value. The size of G value determines the difficulty of buffer design. In order to simplify the design and save cushioning materials, the brittleness value of the product should be improved as much as possible when designing the product. The brittleness value of the product in the software buffer design parameters is determined in Table 1.

The equivalent drop height represents the drop impact energy of the package during circulation. The higher the drop height, the greater the impact energy, the easier the damage to the product, and the higher the buffer protection requirements. Falling mainly occurs in manual loading and unloading and mechanical operation during loading and unloading. The impact energy during falling is mainly considered according to the comprehensive factors such as product weight, volume, personnel height, handling difficulty, and handling mode [21]. The equivalent drop height is usually determined by empirical formula method and standard value method.

3.7. Empirical Formula Method. During manual loading and unloading, the equivalent drop height of packages above 16 kg can adopt the following empirical formula:

$$h = \frac{300}{\sqrt{w}}, \tag{1}$$

where $w$is the weight of package in kg and h is the drop height in cm.

3.8. Standard Value Method. The standard value method is to determine the equivalent drop height according to some national standards and industrial standards of China, the United States, and Japan. The following is a brief introduction to determining the equivalent drop height through the classification of circulation environmental conditions. In the process of circulation, through the links of loading and unloading, handling, transportation, and storage, the conditions are divided into three levels, as shown in Table 2.

First-class circulation conditions are as follows: long transportation distance and many times of transfer, poor loading and unloading conditions, manual loading and unloading difficult to handle with care, etc. Secondary circulation conditions are as follows: first, the transportation distance not long and small number of transfers, when the loading and unloading conditions are better than the primary circulation conditions, and in case of rough loading and unloading conditions and good circulation conditions. According to the circulation condition level of the transportation environment, the equivalent drop height can be determined according to Table 2 in the design of buffer packaging.

In addition, the drop height can also be determined by the corresponding relationship between the weight and size

Table 1: Brittle value of common products.

Trade name	Allowable brittleness value
Large electronic computers, etc.	Below 10
Advanced precision equipment/electronic equipment, etc.	15~20
General electronics, precision equipment	25~40
Computer/general test equipment, etc.	40~60
Advanced watch, clock/advanced color display device	45~65
General household appliances/printing equipment, etc.	60~85
Egg	45~90
Light bulb	50~90
Beer bottle	130~170
Ceramics with complex structure	70~120
General machinery and equipment	Above 110

Table 2: Drop height values.

Weight (kg)	Drop height		
Circulation conditions	Class A	Class A	Class A
<25	90	90	90
26~50	65	65	65
51~75	50	50	50
76~100	45	45	45

of goods, their loading and unloading mode, and the maximum drop height, as shown in Table 3.

Vibration refers to the reciprocating motion of a particle relative to its equilibrium position. The most basic parameters describing vibration are frequency and acceleration. The factors affecting package vibration come from the types of transportation tools, transportation environment, packaging structure, loading weight, and so on. During the circulation of products (i.e., road transportation, railway transportation, air transportation, and sea transportation), they will be subject to vibration, and the vibration impact caused by each transportation mode is also different.

3.9. Automobile Transportation Vibration. The vibration acceleration of automobile transportation is related to the road condition, driving speed, vehicle type, and load capacity. The main factor is the fluctuation and unevenness of the highway.

3.10. Train Transportation Vibration. When the train passes through the rail joint, the wheels are impacted, which is a periodic excitation to the vehicle, which causes the periodic forced vibration of the running vehicle. In normal operation, station entry and exit, crossing, vehicle body shaking, vehicle body vibration, rail joint, bridge girder, and other operations, the vibration caused by rail joint is the strongest, as shown in Table 4.

3.11. Aircraft Transportation Vibration. During air transportation, the vibration of the aircraft mainly comes from the engine vibration, which shows the characteristics of single vibration and high frequency. Its vibration acceleration is small and stable, as shown in Table 5.

3.12. Marine Vibration. The following table shows the vibration data measured at different parts of the goods. It can be seen from the table that the left and right vibration accelerations of the tail deck, front compartment, and rear compartment are the maximum when the cargo ship is sailing [22], as shown in Table 6.

The cushioning packaging material shall be able to transfer the impact and vibration of the external action on the package to the contents after buffering, avoid stress concentration, protect the contents, support the contents, and protect the product structure and cushioning packaging system. Buffer characteristic curve mainly includes maximum acceleration static stress curve, buffer coefficient maximum stress curve, vibration transmission rate frequency characteristic curve, etc.

3.13. Common Cushioning Materials

3.13.1. Foam Plastics. Foam plastics is a material blended with gas and plastics. It is made of foam resins such as PE, PS, and PU. It has excellent cushioning and vibration absorption properties, light weight, easy forming, good protection, high quality and low cost, and being easy to spread. Commonly used foam plastics are EPS and EPE, which are widely used in cushioning packaging of fragile materials such as precision instruments, household appliances, and medical devices. Although foamed plastics have good protective properties and cost performance, they also have disadvantages such as difficulty in recycling, natural weathering, and incineration, which cause harmful gas and environmental pollution. From the perspective of development trend, foam plastics will develop towards being biodegradable and environmentally friendly [23].

3.13.2. Corrugated Board Cushion. Corrugated board has the advantages of good workability, low cost, wide use temperature range, wide plastic foam, and no pollution. But at the same time, there are also some disadvantages: hard surface, inability to directly contact the product when packaging high-grade products, poor moisture resistance, and small recoverability, which is not suitable for the situation of large impact load. From the development trend, corrugated board will develop in the direction of

Table 3: Loading and unloading environment and drop height.

Goods		Loading and unloading mode	Drop parameters	
Weight (kg)	Size (cm)		Attitude	Height (cm)
9	122	One throw	An end face or corner	107
9~23	91	Carried by one person	An end face or corner	91
23~45	122	Two-person handling	An end face or corner	61
45~68	152	Two-person handling	An end face or corner	53
68~90	152	Two-person handling	An end face or corner	46
90~272	183	Mechanical handling	Bottom surface	61
272~1360	Unlimited	Mechanical handling	Bottom surface	46
>1360	Unlimited	Mechanical handling	Bottom surface	30

Table 4: Vibration caused by railway and highway transportation.

Type of transportation	Operation		Maximum acceleration (g)		
			Up and down	Left and right	Front and back
Railway wagon	Vibration during operation (30~ 60 km/h)		0.2~0.6	0.1~0.2	0.1~0.2
	Vibration during deceleration		0.6~1.7	0.2~1.2	0.2~0.5
Automobile	General highway	Good pavement	0.4~0.7	0.1~0.2	0.1~0.2
	20~40 km/h	Bad pavement	1.3~2.4	0.4~1.0	0.5~1.5
	Paved highway	The full load	0.6~ 1.0	0.2~0.5	0.1~0.4
	50~ 100 km/h	No load	1.0~1.6	0.6~1.4	0.2~0.9

Table 5: Vibration generated during air transportation.

Flight status vibration quantity	Vertical (g)	Longitudinal (g)	Transverse (g)	Loading condition
Engine start	0.45	0.1	0.2	No load
Glide	0.38	0.05	0.03	No load
Flight	0.36	0.10	0.04	Load
Turn and circle	0.60	0.08	0.03	Load
Landing	0.40	0.14	0.18	No load

Table 6: Vibration caused by sea transportation.

	Frequency (Hz)			Acceleration (m/s^2)		
	Up and down	Left and right	Front and back	Up and down	Left and right	Front and back
Engine room	17~95	35~128	49.5	0.17~8.2	0.11~3	0.03
Stern deck	21~95	30~132	43~148	0.1~2.8	0.02~2.2	0.02~0.04
Front compartment	21~145	35~153	31~121	0.02~0.25	0.014~2.1	0.003~0.014
Rear compartment	41~207	12~153	45~48	0.01~0.74	0.01~1.7	0.006~0.007

miniaturization, diversified structure, small quantity, low cost, and high quality.

3.13.3. Honeycomb Paperboard Liner. Honeycomb paperboard has a unique inner core structure which is full of air and is not circulated. It has good toughness, resilience, sound insulation, high rigidity, and less consumables. It has the highest unit volume energy absorption value in all cushioning materials, and high thickness honeycomb paperboard can replace plastic foam cushion. Compared with corrugated paperboard, honeycomb paperboard has better pressure bearing performance and bending resistance. The combined application of corrugated paperboard and honeycomb paperboard, combined with their advantages, has better performance. The production of honeycomb paperboard adopts recycled paperboard materials and water-soluble adhesives, which can be completely recycled and avoid the harm to the environment caused by recycling. It is suitable for the transportation and packaging of precision instruments, instruments, household appliances, and fragile products. However, due to the short development time of honeycomb paperboard, the utilization rate of production equipment is still low, and the price of honeycomb paperboard is high. Therefore, improving the production technology of honeycomb paperboard and further reducing the price are the focus of future research.

3.13.4. Pulp Mold. Pulp molded products are made of pulp or wastepaper as the main raw material, which is crushed, pulped, seasoned, molded into various shapes through a drainage metal mesh mold, and then compacted and dried. The product has rich sources of raw materials, pollution-free production and use, light weight, and high compressive strength. Pulp molded products have developed rapidly in China, but so far, the structural design of products still adopts empirical design method, which lacks mature theory and specification. At present, it is only limited to the packaging of small electronic products and eggs.

3.13.5. Plant Fiber Cushioning Material. Plant fiber buffer materials are made of plant fiber (wastepaper and other plant fiber materials) and starch additive materials. At present, the developed materials include foam packing made of straw buffer packaging material, polylactic acid foam packaging material, wastepaper, and starch. It has the characteristics of no environmental pollution, simple production process, low cost, and rich source of raw materials. In the future, the research focus of plant cushioning packaging materials will be the formulation of production technology and the determination of cushioning performance parameters in the actual use of materials.

3.13.6. Expanded Perlite Cushion. Expanded perlite is a white or light colored high-quality thermal insulation material made of acid volcanic glassy lava (perlite) after crushing, screening to a certain particle size, preheating, and instantaneous high-temperature roasting. The inside of the particles is honeycomb structure, nontoxic and tasteless, noncorrosive and noncombustible, and acid and alkali resistant. Different adhesives can be used to make products with different properties. It is characterized by light weight, good heat insulation and sound absorption performance, rich raw materials, low price, safe use, and convenient construction. The expanded perlite cushion liner is made by foaming and molding technology during the bonding process. It is suitable for cushioning packaging of small electronic products and has broad development prospects.

The performance comparison of common cushioning materials is shown in Table 7.

The buffer design is shown in Figure 3.

In practice, the circulation process of products is very complex. There are other requirements for cushion pad besides cushion. Therefore, after the preliminary design of the buffer pad is completed, various performance checks should be carried out to adjust the material and size of the buffer design, and the antivibration check, angle drop check, deflection check, center of gravity check, and creep variable check of the buffer pad should be carried out, respectively.

3.13.7. Antivibration Check. From the initial elastic rate E and bearing area a of the buffer pad, the natural frequency f_o of the buffer package can be obtained. From f_o, the maximum transmission frequency T_m can be obtained from the buffer material database. From the vibration acceleration a in the circulation environment database, the maximum response speed $p = T_m \cdot a$ can be obtained. After void Step4: onshockproof() calculation, if the p value is less than the allowable brittleness value of the product, it is considered to meet the antivibration requirements. If the p value is greater than the allowable brittleness value of the product, redesign it, as shown in Figure 4.

3.13.8. Angle Drop Check. In the actual circulation process, the falling posture of packages can be divided into three types: angular landing (called angular drop), corrugated landing (called corrugated drop), and surface landing (called surface drop). When the package falls, the stress changes greatly, so it is necessary to check and adjust the preliminary design size of the buffer pad. This subject adopts the local buffer packaging method to design and check the angle drop.

$$\frac{3lbh}{\sqrt{l^2 + b^2 + h^2}} > A, \tag{2}$$

where l, b, and h are the length, width, and height of the product, respectively.

The length, width, and height of the product are passed in by the global variable's length, width, and height, and the pad area a is passed in by the global variable area.

After void Step4: onconnerdrop() calculation, assign the pad area to the global variable area.

3.13.9. Deflection Check. When the slenderness ratio (i.e., the ratio of area to thickness) of the pad exceeds a certain limit, the pad is easy to flex or bend, which greatly reduces the load-bearing capacity of the pad. In order to avoid deflection, the ratio of the minimum bearing area a to the thickness t shall comply with the following provisions:

$$A_{\min} > (1.33T)^2. \tag{3}$$

The pad area a and thickness T are passed in through the global variables area and thickness.

After void Step4: 0nflexibility() calculation, assign the pad area to the global variable area.

3.13.10. Creep Check. Under the action of long-term static pressure, the plastic deformation of the buffer material will gradually increase, resulting in the smaller size of the liner and the cracks in the container. This will increase the vibration of the product and the friction with the container and reduce the cushioning capacity of the liner. Therefore, it should be considered to increase the creep compensation value for the pad size in the design.

$$T_c = T(1 + C_T), \tag{4}$$

where $T_C$is the corrected thickness in cm, C_T is the creep coefficient in %, and $T$is the original design thickness.

The liner thickness t is introduced through the global variable thickness, and C_T is taken as a constant of 10%.

After void Step4: onwiggle() calculation, assign the pad thickness to the global variable thickness.

3.13.11. Verification of Center of Gravity. Projected from the center of gravity to the pad, the pad is divided into S_1, S_2, S_3, S_4 according to the bearing area. Weight of each

Table 7: Comparison of characteristics of cushioning packaging materials.

Material science characteristic	Residual deformation	Impact attenuation	Density	Processability		Operating temperature range
				Formability	Adhesion	
Polyethylene foam	Tiny	Excellent	Small	Good	Good	Wide
Polyurethane foam	Tiny	Good	Small	Good	Good	Wide
Styrofoam	Slightly worse	Excellent	Small	Excellent	Good	Narrow
Air cushion plastic film	Difference	Good	Small	Bad	Good	Wide
Foam plastic wire	Varies by material	Good	Small	—	—	Depending on the material
Foam rubber	Nothing	Excellent	Large	Good	Good	Narrow
Wood silk	Difference	Bad	Commonly	—	—	Wide
Corrugated board	Difference	Bad	Commonly	May not	Good	Wide
Crepe paper pad	Difference	Bad	Commonly	May not	Good	Wide
Cellulose acetate silk	Slightly worse	Good	Small	—	—	Slightly narrow
Cellulose acetate silk	Slightly worse	Good	Small	May not	Good	Narrow

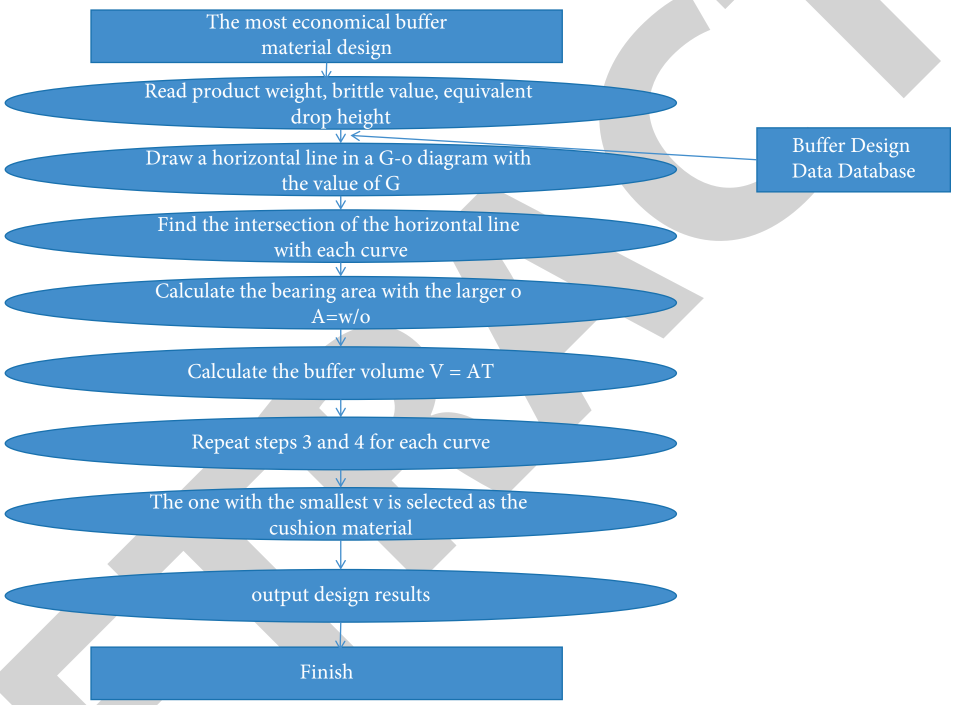


Figure 3: Design flowchart of cushion pad.

pad $W_1 = (S_4/S) \cdot W$, $W_2 = (S_3/S) \cdot W$, $W_3 = (S_2/S) \cdot W$, $W_4 = (S_1/S) \cdot W$. The bearing area of each pad shall be greater than or equal to a. The formula is as follows:

$$A_i = \frac{W_i}{\sigma_{si}}, \tag{5}$$

where A_i is the area of each pad in cm^2 and w_i is the area of each pad, n.

The center of gravity's projection coordinates are transmitted through the global variables coordinate X and coordinate X, the product weight W is transmitted through the global variable weight, and the pad area s is transmitted through the global variable area.

After voidstep4: OnBaryCenter() calculation, assign the pad area of each block to the global variables area, area2, and area3area4, as shown in Figure 5.

4. Experiment and Discussion

Corrugated board is made of corrugated base paper into corrugated shape, and then the surface layer is bonded from both sides with adhesive, so that the center of the board is a hollow structure, which has high strength, stiffness, hardness, pressure resistance, burst resistance, extensibility, and elasticity. The carton made of it is strong and has a wide range of uses [24]. The compressive strength of corrugated board is directly related to the shape of corrugated board. According to the waveform classification seen in the cross section of corrugated board, the corrugated shape is generally divided into U shape, V shape, and UV shape.

The U-shaped corrugated peak is in the shape of circular arc and has a large radius. Corrugated board is elastic, with good extensibility and good reduction performance within

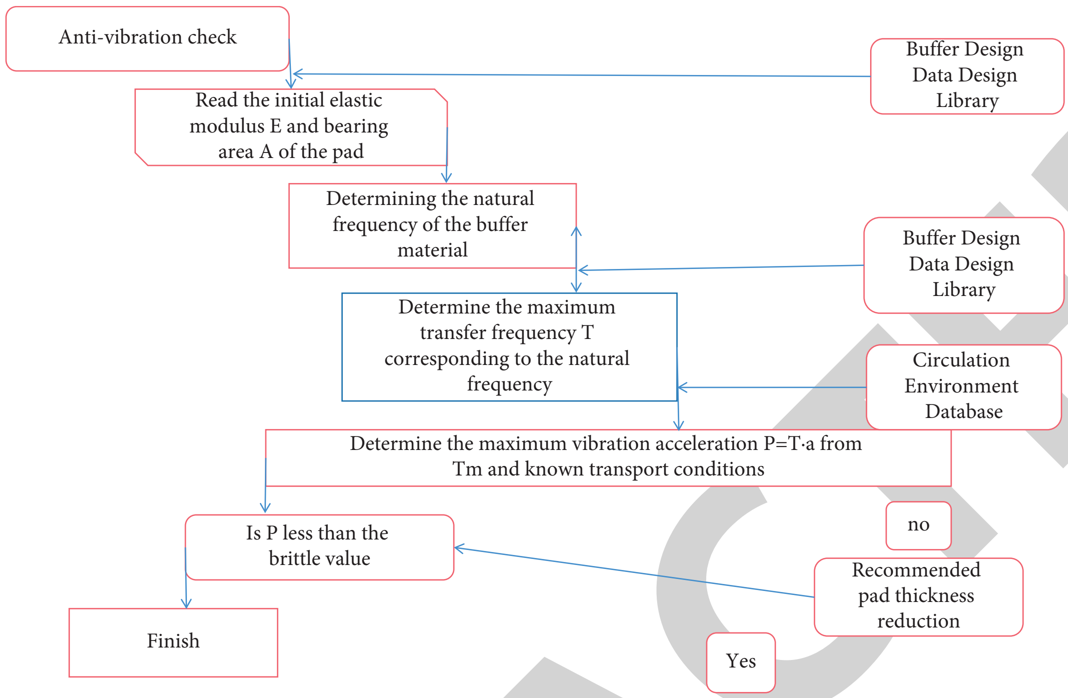


Figure 4: Flowchart of vibration check.

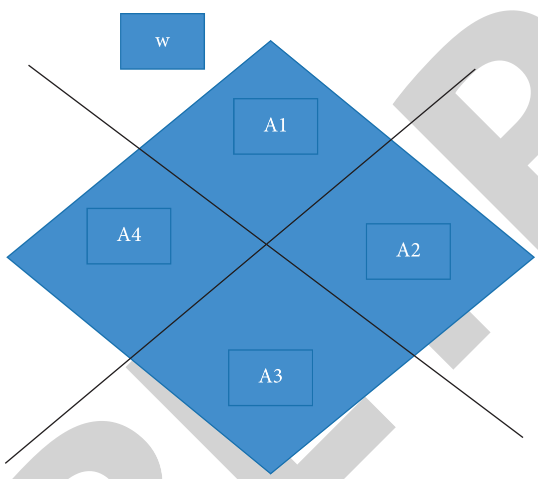


Figure 5: Projection of center of gravity on the bottom.

the elastic limit. It can absorb high energy in the process of compression deformation and has a good cushioning effect. The bonding surface between the top surface of the corrugated board and the surface paperboard is wider than that of the VV shaped paperboard. The amount of adhesive and paper is more, the bonding strength is good, and the wear of the corrugated roller is less. Under the action of large external force, the wave crest of the corrugated core paper is rarely crushed, and most of the straight lines on both sides are bent. The processing performance of U shape is better than that of V shape, but the compressive strength is not high due to the instability of arc force point [25].

V-shaped corrugated board has good stiffness, hardness, and reliability and uses less paper. Because the wave crest radius of V-shaped corrugated board is small and sharp, and the bonding surface between the corrugated top surface and the cardboard surface is narrow, and the amount of adhesive is small, so the bonding strength is also low. During pressing, the corrugated top surface of the core paper is easy to crush and crack, and the corrugating roller wears fast. In practical application, it will be subjected to pressure in three directions, namely, plane pressure, vertical pressure, and parallel pressure. If the V-shaped corrugated board is subjected to plane pressure, the skew is small at the initial stage of pressurization, but it will be damaged when the pressure exceeds the limit point of the board, and the corrugated board cannot return to the initial shape. Due to the poor resilience and poor elasticity of V-shaped corrugated board, it is almost not used now [26].

UV corrugated wave crest is between U-shaped and V-shaped corrugated boards, which combines the advantages of the two. It has high compressive strength, strong bearing capacity, and good bonding strength. When the external force exceeds its bearing capacity, it will not destroy the whole shape of the ridge, and the shape can be basically restored after the external force is eliminated. At present, all kinds of corrugating machines basically use this toothed corrugating roller.

The experimental results show that the three kinds of corrugated boards are subjected to different plane limit pressures, and the deformation degree is V shape, U shape, and UV shape in turn.

The quantity of corrugated board refers to the weight per square meter, expressed in g/m^2. In actual production, you can cut a certain size of corrugated board with a paper cutter and then weigh it. Generally, before production, the theoretical quantity of corrugated board can be estimated according to the raw materials used, that is,

$$G = \sum g_1 + \sum g_m \gamma + \sum g_j + \sum g_a, \quad (6)$$

where G is the quantity of corrugated board in g/m^2, $g_1, g_m, g_j, g_a$are the consumption of a box of board paper, corrugated base paper, laminated paper, and adhesive in g/m^2, and Y is the corrugated coefficient.

By comparing the theoretical quantitative results with the actual measured results, we can analyze the problems existing in the quality and shape of base paper [27].

The thickness of corrugated board is a very important factor in the design of corrugated box, which can be measured by a special corrugated board thickness meter. Its structure is the same as that of ordinary paperboard thickness gauge, but the contact surface with the sample is 10 cm^2 and the pressure is 20 kPa. The estimated thickness can be calculated by the following formula:

$$T = \sum t_1 + \sum t_m, \quad (7)$$

where T is the thickness of corrugated board in cm, t_1 is the thickness of carton board in cm, and t_m is the height of corrugated core paper in cm.

The actual measured thickness of corrugated board is generally less than the calculated value, which shall not be lower than the lower limit of corrugated standard height according to the regulations; otherwise, it is related to the improper control of production process conditions. When designing cartons, the thickness of corrugated cartons does not need to be accurately calculated but can be directly checked in the table.

Corrugated box is a kind of packaging container with a wide range of uses. It is popular because of its unique material structure mechanism, unique packaging effect, and economic benefits. At the same time, it can replace wood and plastic and meet the requirements of environmental protection and low cost. It has a good development prospect.

After finishing the operations such as line pressing, corner cutting, slotting, and printing on the corrugated board, the blank of corrugated box is obtained, as shown in Figure 6.

4.1. Indentation. There are three common indentation types of corrugated boxes: (a) one point type, that is, a single V-groove; the concave surface is a plane, and the punch has a flange. It is mainly used for corrugated folding carton. (b) Three point: this indentation process has two flanges for the female die and one flange for the male die. It is mostly used in the production of double-sided corrugated board and double-core double-sided corrugated board. (c) Five-point type: basically the same as three-point type, but the female die has three flanges and the male die has two flanges. Special for double-core double-sided corrugated board. The above are all rotary indentation processes. They are paired dies that rotate in a disk shape, and the paperboard can leave an indentation between the paired dies. The indentation depth can be changed by adjusting the gap between the female die and the male die.

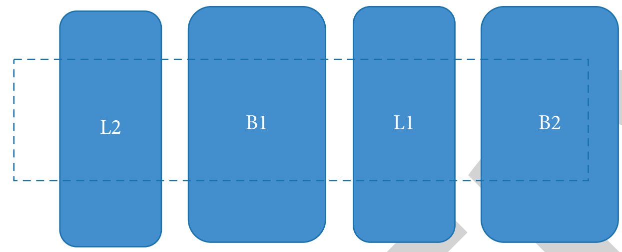


Figure 6: Structure of corrugated box blank.

4.2. Slotting. Slotting refers to cutting a notch on the corrugated board machine that is convenient for folding, and its width is generally the thickness of the board plus 1 mm. The center line of slotting shall be aligned with the center line of indentation as much as possible, and the smaller the deviation, the better. Slotting is closely related to indentation and has an impact on the dimensional accuracy and appearance of cartons.

4.3. Joint. The joint is an essential part for corrugated boxes, especially for type 02 corrugated boxes. When the joints are joined to make cartons, the position of the joints often causes some errors in the internal dimensions of the cartons, so attention must be paid to the design of cartons. The joint mode is the same as the sealing mode. There are three forms: adhesive tape bonding, adhesive bonding, and metal nail bonding.

Based on the inner wall, the effective dimensions in three directions of the box are called the inner dimension of the box.

Factors for Determining Internal Dimensions. The internal dimension is determined by the following factors: the maximum outer diameter of the product and the number of containers of the product; characteristics of packaged goods; considering equipment limitations; arrangement of packages; relevant dimensions of grid lining and buffer in the box.

Calculation formula of internal dimension is as follows:

$$x_i = x_{\max} + T + K, \quad (8)$$

where x_i is the inner dimension of carton, divided into three directions of length, width, and height, in cm, x_i is the inner dimension of carton, divided into three directions of length, width, and height, in cm, $x_{\max}$is the maximum external dimension of the contents, divided into three directions of length, width, and height, in cm, T is the thickness of buffer pad in cm, and K' is the internal dimension correction factor in cm.

The maximum external dimension parameters (length, width, and height) of the contents are input from the global variable's length, width, and height, the thickness of the buffer pad is obtained from the global variable thickness, and the internal dimension correction coefficient is taken from Table 8.

Table 8: Correction factor of inner dimension of corrugated box/cm.

L_i	B_i	H_i		
		Small box	Medium box	Large box
0.3~0.7	0.3~0.7	0.1~0.3	0.3~0.4	0.5~0.7

After void Step4: oncartoninside() calculation, assign L_i (long) to global variables L_1, L_2. Assign B_i (width) to global variable B_1, B_2 and H_i (height) to global variable H.

The manufacturing size of corrugated box is the blanking size when making box. Its value should be greater than the inner diameter dimension, which is the basis of calculation and design dimension. Generally, when measuring the size of cartons, the products are packed and tied tightly and then amplified according to the tolerance coefficient to ensure that they are not too tight during packing. When making boxes, the manufacturing and cutting dimensions are calculated according to the inner diameter specification and considering the extension value. Therefore, the dimensions of corrugated boxes are subject to the measured outer diameter, so the actual outer diameter must be printed when printing the box surface of corrugated boxes. The theoretical value of the manufacturing size of corrugated box is equal to the inner diameter size plus several times the thickness of corrugated board (total thickness) and the shrinkage caused by indentation (called enlarged value), among which the thickness of corrugated board and the size of shaking cover structure play an important role. The calculation formula is

$$x = x_i + k, \tag{9}$$

where x is the manufacturing dimension of corrugated box, divided into five directions: I, L_2, B, B_2, and h, in cm, X_i is the inner dimension of carton, divided into three directions of length, width, and height, in cm, and K is the correction coefficient of inner dimension of carton manufacturing, in cm.

The dimension parameters (length, width, and height) in the carton are imported from the global variables L_1, B_1, H, the dimension correction coefficient K in carton manufacturing is taken from Figure 7, and the corrugated shape is imported from the global variable corrugating, as shown in Table 9.

In the case of butt sealing of carton swing cover, such as 0201, 0204, and 0207, the theoretical value of manufacturing dimension of swing cover width shall be 1/2 of the manufacturing dimension of carton width. However, due to the impact of the rebound effect of the rocking cover, a gap must be generated at the butt joint of the rocking cover, so that the box is not tightly sealed, resulting in dust pollution of the contents, as shown in Figure 8.

When storing and transporting goods packed in corrugated boxes, the calculation method used and the volume of the marked box printed on the box surface are based on the outer diameter. Therefore, the outer diameter of the corrugated box should be calculated in the design. Generally, the size of corrugated box is determined according to the inner diameter size, and the outer diameter size is calculated according to the manufacturing size, as shown in Table 9.

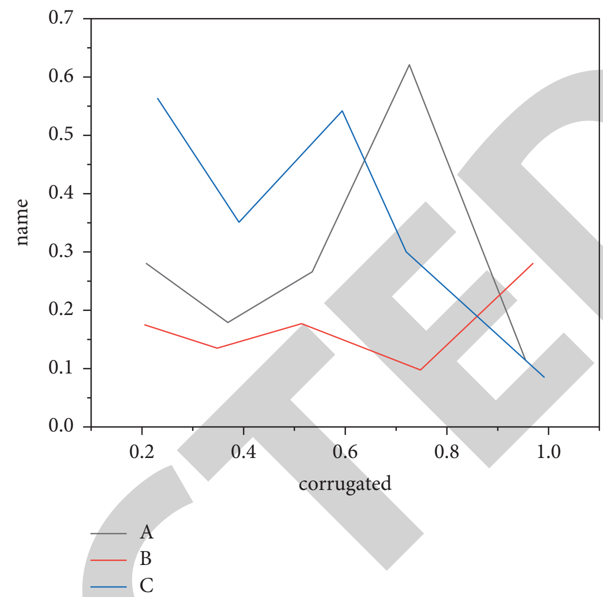


Figure 7: Correction factor of manufacturing dimension of corrugated box/cm.

Table 9: Correction factor of external dimension of corrugated box/cm.

Ridge type	A	B	AB
K	0.5~0.7	0.3~0.5	0.8~1.2

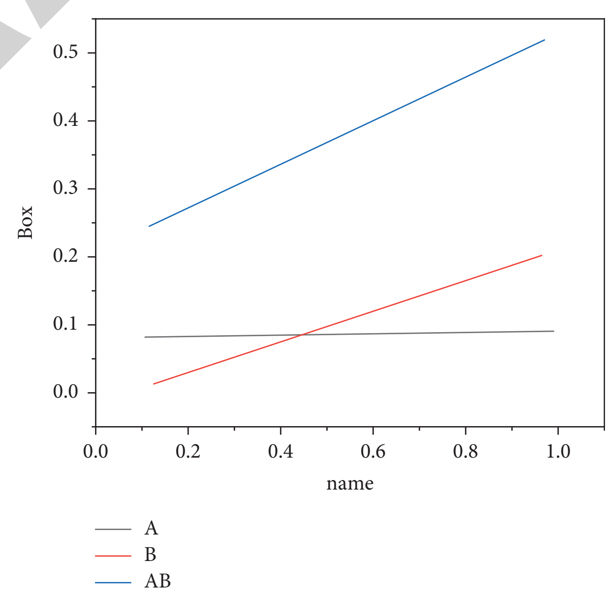


Figure 8: Elongation coefficient of type 8 carton flap/cm.

After void Step4: oncartoninside() calculation, assign L_i (long) to global variables L_1, L_2. Assign B_i (width) to global variable B_1, B_2 and H_i (height) to global variable H.

According to the calculation formula of corrugated box compressive strength, the necessary corrugated box strength

Table 10: Thickness of corrugated board.

Ridge type	Paperboard thickness
A	5.3
B	3.3
AB	8.1

can be calculated according to the predetermined conditions to see whether it meets the requirements. On the contrary, a certain corrugated board can also be selected according to the predetermined strength, and then a certain corrugated board base paper can be selected. The formula of compressive strength can be divided into two categories: one is calculated according to the test strength of corrugated board base paper, i.e., face paper and core paper. The other is calculated directly according to the test strength of corrugated board, as shown in Table 10.

5. Conclusion

The development of corrugated box structure CAD system adopts the software engineering method and compiles the program with VC++ 6.0. The main characteristics are as follows: the popular and mature Windows 98 is used as the operation platform of the system, which gives full play to the characteristics of easy use and operation of windows graphical interface. It has low requirements for users and can operate skillfully without mastering a lot of professional knowledge. Visual C++ 6.0 is used as the software development tool, and the GUI (graphical device interface) features provided by various tools and other high-level languages are used to design a simple and friendly user interface. MFC application framework is used to create applications. MFC application framework integrates some basic functions required by applications and carries out software development and function expansion on this basis. Using the basic concepts and characteristics (encapsulation, inheritance, and polymorphism) of classes and objects in C++, a user-defined class is established. Establish the data structure (data and operation) of the graphic elements required for the corrugated box plane expansion drawing; that is, establish the basic graphic element classes, including straight line class, rectangle class, arc class, ellipse class, free curve class, and text class. On this basis, establish the data model of the plane expansion drawing and draw it.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the 2021 Xi'an Polytechnic University Teaching Reform Project: the practice research of packaging design curriculum system optimization and teaching content reform based on "innovative talent training objective" (no. 21JGYB18).

References

[1] K. Svidlo and T. Gavrish, "Scientific justification of custard semi-finished product technology of herodietic purpose," *Restaurant and hotel consulting. Innovations*, vol. 3, no. 2, pp. 210–220, 2020.

[2] K. Svidlo, A. Sobko, L. Karpenko, and T. Gavrish, "Improvement of the technology of shortcrust baked semi-finished product on the basis of model functional compositions," *Eastern-European Journal of Enterprise Technologies*, vol. 2, pp. 61–67, 2021.

[3] N. V. Linovskaya, E. V. Mazukabzova, E. V. Mazukabzova, N. B. Kondratev, and E. N. Krylova, "The study of the technological adequacy of raw materials used in the production of chocolate semi-finished product," *Vestnik MGTU*, vol. 22, no. 3, pp. 404–412, 2019.

[4] Z. Meng, "Research on the application of digital technology in product packaging design," *Paper Asia*, vol. 2, no. 3, pp. 144–148, 2019.

[5] W. Yu and P. Sinigh, "Application of cad in product packaging design based on green concept," *Computer-Aided Design and Applications*, vol. 19, no. 2, pp. 124–133, 2021.

[6] S. B. Sidelnikov, O. V. Yakivyuk, V. N. Baranov, E. Y. Zenkin, and I. N. Dovzhenko, "Development, modeling and research of technology for producing longish deformed semi-finished products from aluminum-magnesium alloys with low scandium contents," *Izvestiya Vuzov. Tsvetnaya Metallurgiya (Universities' Proceedings Non-Ferrous Metallurgy)*, no. 6, pp. 51–59, 2019.

[7] M. Deja, M. Dobrzyński, and M. Rymkiewicz, "Application of reverse engineering technology in part design for shipbuilding industry," *Polish Maritime Research*, vol. 26, no. 2, pp. 126–133, 2019.

[8] N. N. Tipsina, N. G. Batura, E. L. Demidov, and M. S. Beloshapkin, "The use of soy in the production of nutrition products and the prospects of the development of application of soy semi-finished productsin bakery production," *Bulletin of KSAU*, no. 1, pp. 163–168, 2021.

[9] O. Cherevko, V. Mykhaylov, A. Zahorulko, A. Zagorulko, and I. Gordienko, "Structural and mechanical properties of the developed fruit and berry semi-finished product," *Eureka: Life Sciences*, no. 1, pp. 25–30, 2021.

[10] G. A. Bakhadirov and K. Khusanov, "Construction of automatic lines connecting several mechanical operations in processing leather semi-finished product," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 4, pp. 2079–2090, 2020.

[11] O. Sahmel, S. Siewert, W. Schmidt, K.-P. Schmitz, and N. Grabow, "Cross-sectional analysis of tubular polymer semi-finished products using ultrasound in comparison with other measuring methods," *Current Directions in Biomedical Engineering*, vol. 6, no. 3, pp. 151–154, 2020.

[12] N. Golembovskaya, "Usage of chia seeds in the composition of dietary semi-finished minced prod-UCTs," *Scientific Messenger of LNU of Veterinary Medicine and Biotechnologies*, vol. 21, no. 92, pp. 19–22, 2019.

[13] N. Hrudkina, L. Aliieva, P. Abhari, M. Kuznetsov, and S. Shevtsov, "Derivation of engineering formulas in order to calculate energy-power parameters and a shape change in a semi-finished product in the process of combined extrusion,"

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9823467, 1 page
https://doi.org/10.1155/2023/9823467

Retraction

Retracted: Correlation Analysis of Interbank Money Market Interest Rate and Financial Crisis Based on Neural Network Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Zeng, C. Yuan, and C. Liang, "Correlation Analysis of Interbank Money Market Interest Rate and Financial Crisis Based on Neural Network Model," *Security and Communication Networks*, vol. 2022, Article ID 8807215, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8807215, 11 pages
https://doi.org/10.1155/2022/8807215

Research Article

Correlation Analysis of Interbank Money Market Interest Rate and Financial Crisis Based on Neural Network Model

Xiaohua Zeng,[1] **Chiping Yuan**,[2] **and Changzhou Liang**[1]

[1]*School of Economics and Trade, Guangzhou Xinhua University, Dongguan 523133, China*
[2]*Lingnan College Sun Yat-Sen Universtiy, Guangzhou 510275, China*

Correspondence should be addressed to Chiping Yuan; yuanchip@sysu.edu.cn and Changzhou Liang; salcz@xhsysu.edu.cn

Received 17 March 2022; Revised 13 April 2022; Accepted 21 April 2022; Published 27 May 2022

Academic Editor: Chin-Ling Chen

This paper combines the neural network model to analyze the correlation between the interbank money market interest rate and the financial crisis, and theoretically analyzes the influence of the characteristics of the banking industry on the transmission of monetary policy interest rates. Moreover, this paper constructs a theoretical model to analyze the influence of banking industry characteristics on the transmission of monetary policy interest rate and examines whether the variables of banking industry characteristics still affect the measurement indicators of monetary policy interest rate transmission after the financial crisis subsides. In addition, this paper combines the neural network model to construct a correlation analysis system between the interbank money market interest rate and the financial crisis. The experimental research shows that the correlation analysis between the interbank money market interest rate and the financial crisis based on the neural network model proposed in this paper can play a certain role.

1. Introduction

Under normal circumstances, the marketization of interest rates will directly affect the risk management ability of commercial banks and improve the pricing ability of commercial banks on deposit and loan interest rates, and there is a relatively low interest rate risk. After the interest rate liberalization, in the short term, the interest rate of bank deposits may rise and the interest spread will narrow. Moreover, it may also lead to large fluctuations in financial asset prices in the short term, intensify competition among banks, and lead to risks such as strengthening the trend of integrated operations of financial institutions. The relationship between interest rate liberalization and macroeconomics is uncertain. Existing research is generally divided into stages of development. In developed countries, interest rate liberalization will help increase macroeconomic output under normal conditions. However, its occurrence in developing countries increases the volatility of macroeconomic output.

Interest rate marketization refers to the fact that the central bank gives the right to determine interest rates to financial institutions, which independently determine interest rates in accordance with the supply and demand of market funds. At the same time, the central bank indirectly regulates interest rates by controlling benchmark interest rates, deposit reserves, re-lending rates, re-discount rates, open market operations, and other means to transmit national policies. Interest rate liberalization is beneficial to the improvement and optimization of the financial environment, and can help improve the efficiency of operating the financial system and play a potential role in economic development. In particular, for developing countries, the microeconomic foundation and macroeconomic environment are weak, and the risk of interest rate liberalization reform is greater and more complicated. Therefore, it is of great significance to explore and prevent financial risks such as liquidity risk, credit risk, and exchange rate risk in the process of interest rate marketization.

Interest rate liberalization increases the autonomy of commercial banks and eliminates the market where commercial banks strictly enforce interest rates prescribed by the People's Bank of China, distort price relations, lower interest rates to optimize financial services, and violate "quality-based pricing" and "risk-reward matching" phenomenon of the principle resource allocation function. Commercial banks may, based on bank operation and management and market interest rates, comprehensively consider their own operating costs, customer risk differences, target returns, industry competition, and the elasticity of interest rate supply and demand, determine the interest rate levels for deposits and loans, and implement differentiated strategies. The liberalization of interest rate has prompted commercial banks to develop intermediary business and carry out product innovation. At present, the ratio of intermediary business to the income of commercial banks is still low. The marketization of interest rate has greatly reduced the interest spread income of commercial banks in deposits and loans, and the profit point of state-owned commercial banks has turned more to the development of intermediary business. The impact of the financial crisis has also brought about an internationalized financial service concept, and the increasingly mature market has an increasing demand for financial services, which greatly expands the profit margins of commercial banks for intermediary business. The liberalization of interest rates enables commercial banks to have more means to actively match assets and liabilities. Usually, commercial banks cannot match assets, liabilities in terms of maturity, and interest rates. However, after the interest rate marketization, commercial banks can reduce or increase interest rates through active negative detection methods to absorb the deposits required by commercial banks and ultimately reduce the risk of mismatch. Also, under the increasingly perfect financial market, commercial banks use financial derivatives to alleviate the risk of asset-liability gap, reduce the possibility of loss of benefits, and better manage assets and liabilities of commercial banks.

Based on the above analysis, this paper combines the neural network model to analyze the correlation between the interbank money market interest rate and the financial crisis, which provides a method for the discovery of the risk factors of the financial crisis and improves the effective response strategy of the financial crisis.

2. Related Work

Literature [1] focuses on the analysis of optimal fiscal policy and monetary policy; Literature [2] studies the optimal monetary policy rules; Literature [3] studies the optimal monetary policy under sticky prices; Literature [4] studies a comprehensive effect assessment of the macro model of monetary policy which is presented. There are more and more macro studies that add financial factors. In recent years, the DSGE model has been widely used in the study of financial crisis. Literature [5] discusses the selection and application of optimal monetary policy rules and establishes an open economy DSGE model with a "financial accelerator," which lays the early foundation for the application of monetary policy rules. Literature [6] constructs a new Keynesian DSGE model from different channels and compares the optimal monetary policy choices of different parameters of economic fluctuations under Taylor's rule; Literature [7] discusses whether monetary policy is a quantitative rule or a price rule, and argues that the interest rate rules will become an important and effective means of monetary policy; Literature [8] adds the banking sector to the general equilibrium model, examines the relationship between bank credit and economic fluctuations, and finds that the impact of banks on economic fluctuations has a strong explanatory power; Literature [9] studied monetary policy options from different perspectives, and they all came to the conclusion that it is necessary to strengthen the coordination and cooperation between policy tools, and it is necessary to use mixed policy tools; and it is necessary to use mixed policy tools; Literature [10] studied the impact of monetary policy on bank risk and believed that monetary policy and macro-prudential policies need to be coordinated with each other; Literature [11] studied the impact of monetary policy on the fluctuation of economic-related variables and believed that the open frame model is the best choice for national interest rate policy. Literature [12] analyzed the effectiveness of monetary policy from different perspectives and methods.

Literature [13] explained the promotion effect of interest rate marketization from the perspective of endogenous growth theory. Interest rate marketization will make savings more effectively allocated to high-yield investment projects. Literature [14] found that the actual average annual economic growth rate of each country will increase by 1%–2% after interest rate liberalization through panel data research on 21 interest rate liberalization countries. However, interest rate liberalization is also a double-edged sword. While bringing benefits, it will also increase risks in the financial sector, aggravate macroeconomic fluctuations, and increase the probability of a financial crisis. Literature [15] analyzed the incomplete statistics of the banking crisis in the interest rate liberalization countries and concluded that the crisis after the interest rate liberalization is not an accidental event, including the developed economies such as the USA and Japan with sound financial infrastructure. It also includes developing countries such as Thailand, Kenya, and Chile with weak financial infrastructure. A large number of studies on banking crises and currency crises show that financial liberalization represented by deregulation of interest rates often leads to problems such as excessive credit, increasing the instability of the financial system, and thus triggering financial crises. Literature [16] used 0/1 dummy variables as proxy variables for interest rate control periods and interest rate relaxation periods. The study found a strong positive relationship between interest rate liberalization and the probability of a subsequent banking crisis. Literature [17] found that banking crises are more likely to occur after the relaxation of interest rate controls and direct credit scale controls. Literature [18] further subdivided financial liberalization into internal financial liberalization represented by interest rate liberalization and external financial liberalization represented by capital account deregulation. Financial liberalization is the main cause of the financial crisis. In general, financial crises are more likely to occur in an open financial system, and the probability of bank crises and currency crises will increase significantly after interest rate liberalization. And the banking crisis has shown a trend following the reform of interest rate liberalization [19].

3. Theoretical Model of the Influence of Banking Sector Characteristics on the Transmission of Monetary Policy Interest Rates

The model assumes that the banking system is an imperfectly competitive market, so banks have the right to set prices in the deposit and loan markets, and can set the loan interest rate i_L and deposit interest rate i_D by themselves, while the government bond interest rate i_B is beyond the control of the bank. The model uses i_B as the monetary policy rate. The assets of commercial bank j include loans issued (L_j), government bonds held (B_j), and reserves (R_j), while liabilities include deposits received (D_j), and there is the following relationship between them:

$$B_j = D_j - L_j - R_j. \tag{1}$$

We assume that the loan demand function and deposit supply function faced by banks are

$$\begin{aligned} L_j &= L\left(i_{Lj} - i_L, i_L\right) = L_0 e \times p - \left[\theta_1\left(i_{Lj} - i_L\right) - \theta_2 i_L\right], \\ D_j &= D\left(i_{Dj} - i_D, i_D\right) = D_0 e \times p\gamma\left[\left(i_{Dj} - i_D\right) + \eta_2 i_D\right]. \end{aligned} \tag{2}$$

Among them, L_0 and D_0 are constants greater than 0, i_L and i_D represent the average loan interest rate and average deposit interest rate in the market, respectively, and i_{Lj} and i_{Dj} are the loan interest rate and deposit interest rate selected by bank j. θ_1 and η_1 are constants greater than 0, representing the semi-elasticity of the bank's loan demand and deposit supply, respectively, with respect to the gap between the bank's loan interest rate and deposit interest rate and the market average. The size of θ_1 is related to the degree of competition in the credit market, and the more intense the market competition, the larger the θ_1. Similarly, θ_2 and η_2 are also constants greater than 0, representing the semi-elasticity of the bank's loan demand and deposit supply with respect to the average loan interest rate and the average deposit interest rate, respectively. If all banks in the market are assumed to be identical, then L_0, D_0 and all semi-elasticities are the same for each bank.

Due to the information asymmetry in the loan market and the existence of contract execution costs (negotiation costs, the cost of investigating the borrower's solvency, and credit history before the transaction, the cost of monitoring the use of the borrower's funds after the transaction), issuing loans has become a high-cost activity. In the case of information asymmetry in the credit market, if the scale of bank loan business exceeds the customer group they trust, the cost of lending will increase significantly. In addition, the larger the loan scale expands, the faster the cost increases; that is, when the credit scale exceeds a certain amount, the bank's loan cost function C_j is a single-increasing convex function, and a quadratic function is used to describe this relationship, namely,

$$C_j = \begin{cases} \gamma_0 L_j, & L_j \le L^*, \\ \gamma_0 L_j + \left(\dfrac{\gamma_1}{2}\right)\left(L_j - L^*\right)^2, & L_j > L^*. \end{cases} \tag{3}$$

L^* represents the maximum loan amount that banks can lend to enterprises with excellent qualifications. If the scale of credit is to be further increased, it can only lend to enterprises with lower credit levels. The model assumes that L^* and γ_0, γ_1 are the same for all banks. The parameter γ_1 plays an important role in the model. The worse the environment in which the bank conducts business (macroeconomic environment, banking competition environment, etc.), the larger the γ_1 will be, and the steeper the cost function will be.

Finally, we assume that banks hold reserves at a fixed ratio, that is, $R_j = \rho D_j$.

The goal of commercial banks is to set the optimal $iL, i\ D$ to maximize their own profits, that is, to solve the following problems under the above assumptions:

$$\begin{aligned} \mathrm{Max}\left(i_{Lj}, i_{Dj}\right) &= i_{Lj} L\left(i_{Lj}, \Lambda\right) + i_B B_j \\ &\quad - c\left[L\left(i_{Lj}, \Lambda\right)\right] - i_{Dj} D\left(i_{Dj}, \Lambda\right) \\ &= i_{Lj} L\left(i_{Lj}, \Lambda\right) + i_B\left[(1-\rho) D\left(i_{Dj}, \Lambda\right) - L\left(L_{Lj}, \Lambda\right)\right] \\ &\quad - c\left[L\left(i_{Lj}, \Lambda\right)\right] - i_{Dj} D\left(i_{Dj}, \Lambda\right). \end{aligned} \tag{4}$$

The first-order conditions for a bank to maximize profits are

$$\begin{aligned} L_j + \left(i_{Lj} - i_B\right) L' - C' L' &= 0, \\ -D_j + \left[i_B(1-\rho) - i_{Dj}\right] D' &= 0. \end{aligned} \tag{5}$$

The optimal loan interest rate is $i_{Lj} = (1/\theta) + \ (i_B + \omega + \gamma_1 L_j)$.

Among them, $\omega = \gamma_0 - \gamma_1 L^*$ is a constant. What we are most concerned about is the effect of the transmission of policy interest rate changes to the loan interest rate, so according to the expression of the optimal loan interest rate, the transmission coefficient $\partial i_{Lj}/\partial i_B$ of the policy interest rate to the loan interest rate is further deduced.

$$0 < \frac{\partial i_{Lj}}{\partial i_B} = \frac{1}{\left(1 + \gamma_1 \theta_1 L_j\right)} < 1. \tag{6}$$

In order to investigate the influencing factors of the transmission coefficient of the policy interest rate to the loan interest rate, the partial derivatives of γ_1 and θ_1 are obtained by using the interest rate transmission coefficient to obtain

$$\begin{aligned} \frac{\partial^2 i_{Lj}}{\partial i_B \partial \gamma_1} &= \frac{-\theta_1 L_j}{\left(1 + \gamma_1 \theta_1 L_j\right)^2} < 0, \\ \frac{\partial^2 i_{Lj}}{\partial i_B \partial \theta_1} &= \frac{-\left(1 - \theta_1 i_L\right)\gamma_1 L_j}{\left(1 + \gamma_1 \theta_1 L_j\right)^2} > 0. \end{aligned} \tag{7}$$

It can be seen that the larger the γ_1 is, the smaller the transmission coefficient of the policy interest rate to the loan interest rate is, and the larger the θ_1 is, the larger the transmission coefficient of the policy interest rate to the loan interest rate is. γ_1 represents the marginal cost of the bank in expanding the scale of credit. A country's macroeconomic environment, financial system environment, and the quality of the banking industry's own assets may all affect the magnitude of this coefficient, and the lower the asset quality of the banking industry. The worse the domestic financial system environment and the larger the γ_1, the weaker the transmission of monetary policy interest rate to loan interest rate. Similarly, we can see from the results that the larger the θ_1, the larger the transmission coefficient from the policy rate to the lending rate. θ_1 represents the semi-elasticity of the loan demand faced by banks with respect to the gap between bank loan interest rates and the market average. When the degree of competition in the banking industry increases, θ_1 increases, and the transmission of policy interest rates to loan interest rates will also increase. To sum up, the transmission effect of monetary policy interest rate to loan interest rate is related to the degree of competition in the banking industry, it is also related to the marginal cost when banks expand the scale of credit, and it is related to the quality of the banking industry's own assets.

In order to study the influence of banking industry characteristics on the transmission effect of monetary policy interest rate to bank loan interest rate, the first problem to be solved is what method to use to measure the transmission effect of monetary policy interest rate to bank loan interest rate. By reviewing previous studies, we found that different scholars have different approaches to this issue. In order to enrich the research content and enhance the credibility of the results, this paper will use the error correction model and the SVAR model to measure the transmission effect of monetary policy interest rates to bank loan interest rates, and the error correction model is used in this section.

An error correction model (ECM) was established for each sample country's monetary policy rate and bank lending rate using the E-G two-step method. The advantage of this approach is that it can separately describe the short-term adjustment relationship and long-term dynamic equilibrium relationship between the two interest rates, and then comprehensively consider the short-term and long-term effects of interest rate transmission. The specific method is to carry out unit root test and cointegration test for the monetary policy interest rate index series $PR_{i,t}$ and bank loan interest rate index series $LR_{i,t}$ of each sample country, and the countries that fail the cointegration test between the two series are deleted from the sample. For each sample country that satisfies the cointegration relationship, a cointegration regression model is established:

$$LR_{i,t} = c_i + \beta_i \cdot PR_{i,t} + \varepsilon_{i,t}. \tag{8}$$

Among them, $LR_{i,t}$ and $PR_{i,t}$ are the bank loan interest rate and the monetary policy interest rate of the sample country i, respectively, and β_i is the long-term transmission coefficient of the change of the monetary policy interest rate in the country to the loan interest rate, which represents the long-term equilibrium relationship between the policy interest rate and the bank loan interest rate. The second step uses the residuals of the first step, the monetary policy interest rate, and the first-order difference term and the lag term of the bank loan interest rate to establish an error correction model, and the short-term parameters are estimated by the following equations.

$$\begin{aligned}\Delta LR_{i,t} = {} & \gamma_i + \alpha_{i1}\Delta LR_{i,-1} + \alpha_{i2}\Delta LR_{i,t-2} + \theta_i \Delta PR_{i,t} \\ & + \theta_{i1}\Delta PR_{i,t-1} + \theta_{i2}\Delta PR_{i,t-2} + \delta_i \hat{\varepsilon}_{i,t-1} + \mu_{i,t}.\end{aligned} \tag{9}$$

Among them, the current influence coefficient θ_i of monetary policy interest rate change $\triangle PR_{i,t}$ on loan interest rate change $\triangle LR_{i,t}$ is the short-term transmission coefficient of policy interest rate change to loan interest rate. The coefficient δ_i before the error correction term $\hat{c}_{i,-1}$ represents the adjustment speed when the two interest rates deviate from the long-term equilibrium relationship. By comparing the long-term interest rate transmission coefficient β_i, the short-term interest rate transmission coefficient θ_i, and the adjustment speed δ_i in different countries, we can compare the differences in the transmission effect of monetary policy interest rates to bank lending rates in different countries.

From the two aspects of the competition degree of the banking system and the asset quality of the banking system, the method of multiple regression is used to examine the influence of the characteristics of the banking industry on the estimated coefficient $\beta_i, \theta_i, \delta_i$ in the model. Among them, regarding the degree of competition in the banking system, this paper uses the market concentration index, that is, the proportion of the total assets of the three largest commercial banks in a country to the total assets of the banking industry (CR3) to reflect. This paper introduces per capita GDP and CPI. Finally, in order to control the factors of the institutional environment, this paper selects the regulatory quality (rq) index measured by the World Bank as the control variable. The regulatory quality index aims to reflect the ability of a government to formulate and implement reasonable policies. Therefore, we use the interest rate transmission coefficient index $\beta_i, \theta_i, \delta_i$ as the explained variable, and the various indicators and other variables of the banking industry characteristics as the explanatory variables to establish the following multiple regression model, and focus on examining the sign of the coefficient of the banking industry characteristic variable and its economic significance.

$$\begin{aligned}\beta_i &= \gamma_0 + \gamma_1 CR3_i + \gamma_2 npl_i + \gamma_3 fd_i + \gamma_4 G\,DP_i + \gamma_5 CPI_i + \gamma_6 rq_i + \varepsilon_1, \\ \theta_i &= \rho_0 + \rho_1 CR3_i + \rho_2 npl_i + \rho_3 fd_i + \rho_4 G\,DP_i + \rho_5 CPI_i + \rho_6 rq_i + \varepsilon_2, \\ \delta_i &= \varphi_0 + \varphi_1 CR3_i + \varphi_2 npl_i + \varphi_3 fd_i + \varphi_4 G\,DP_i + \varphi_5 CPI_i + \varphi_6 rq_i + \varepsilon_3.\end{aligned} \tag{10}$$

The SVAR model is a measurement method that introduces the structural relationship between variables based on economic and financial theory on the basis of the traditional VAR model. The model adds the contemporaneous influence relationship between variables, so as to fully

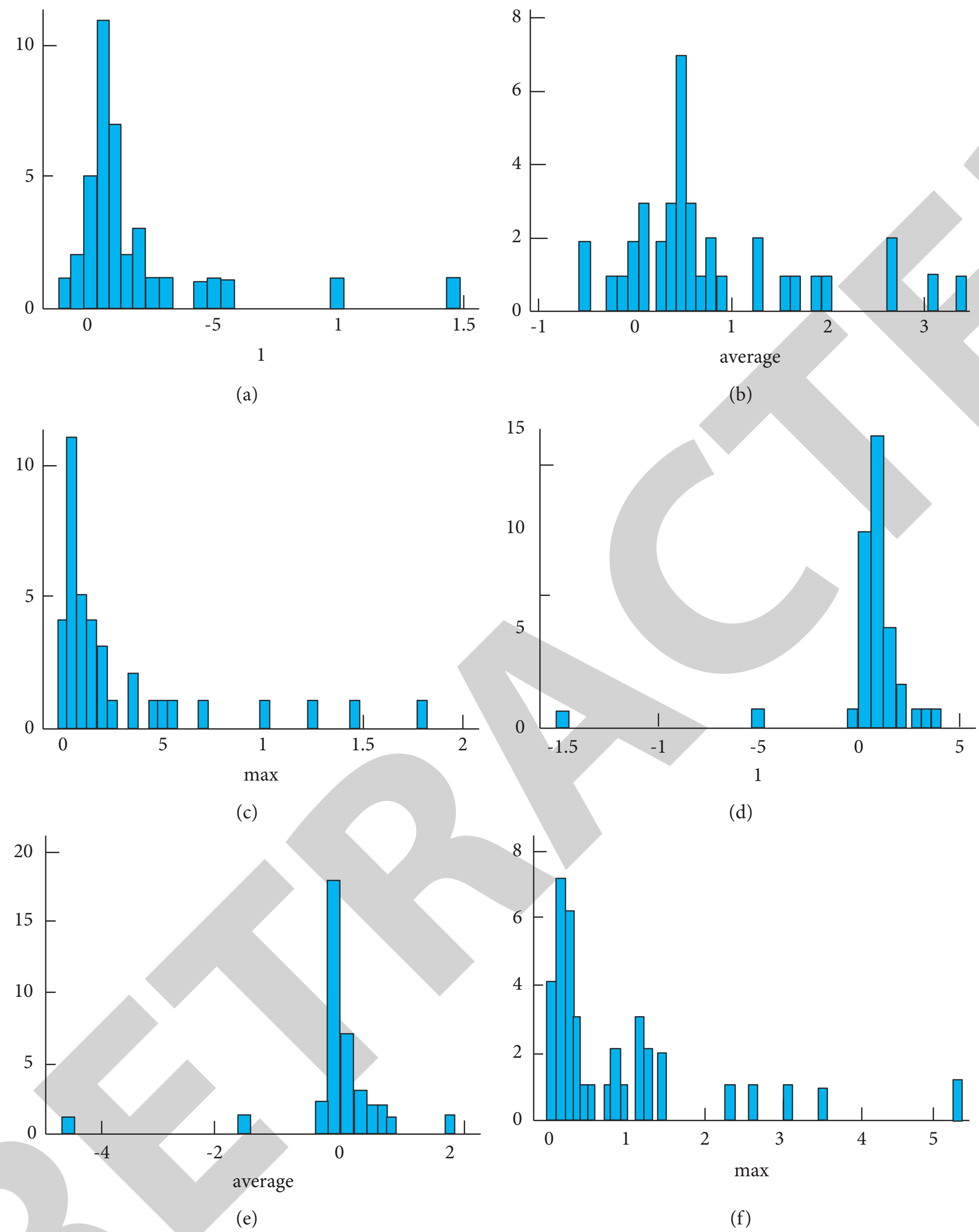
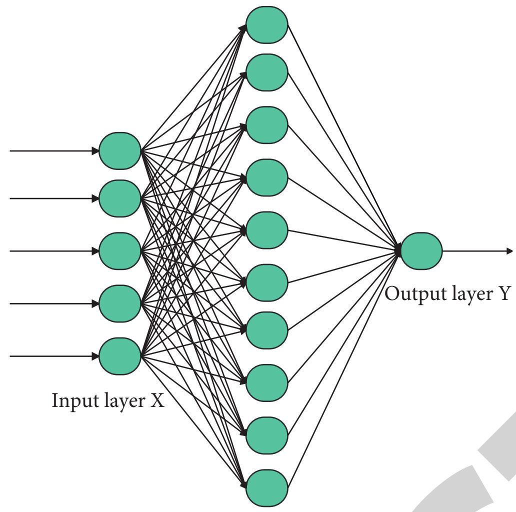


Figure 1: Frequency distribution histogram. (a) Histogram of the frequency distribution of re_1 before the financial crisis. (b) Histogram of the frequency distribution of re_average before the financial crisis. (c) Histogram of the frequency distribution of re_max before the financial crisis. (d) Histogram of the frequency distribution of re_1 after the financial crisis. (e) Histogram of the frequency distribution of re_average before the financial crisis. (f) Histogram of the frequency distribution of re_max before the financial crisis.

consider the influence of its own lag item and other variables and their lag items in the current period. The advantage is that it can capture the immediate structural relationship of variables in the system, which is more in line with the actual situation of economic operation.

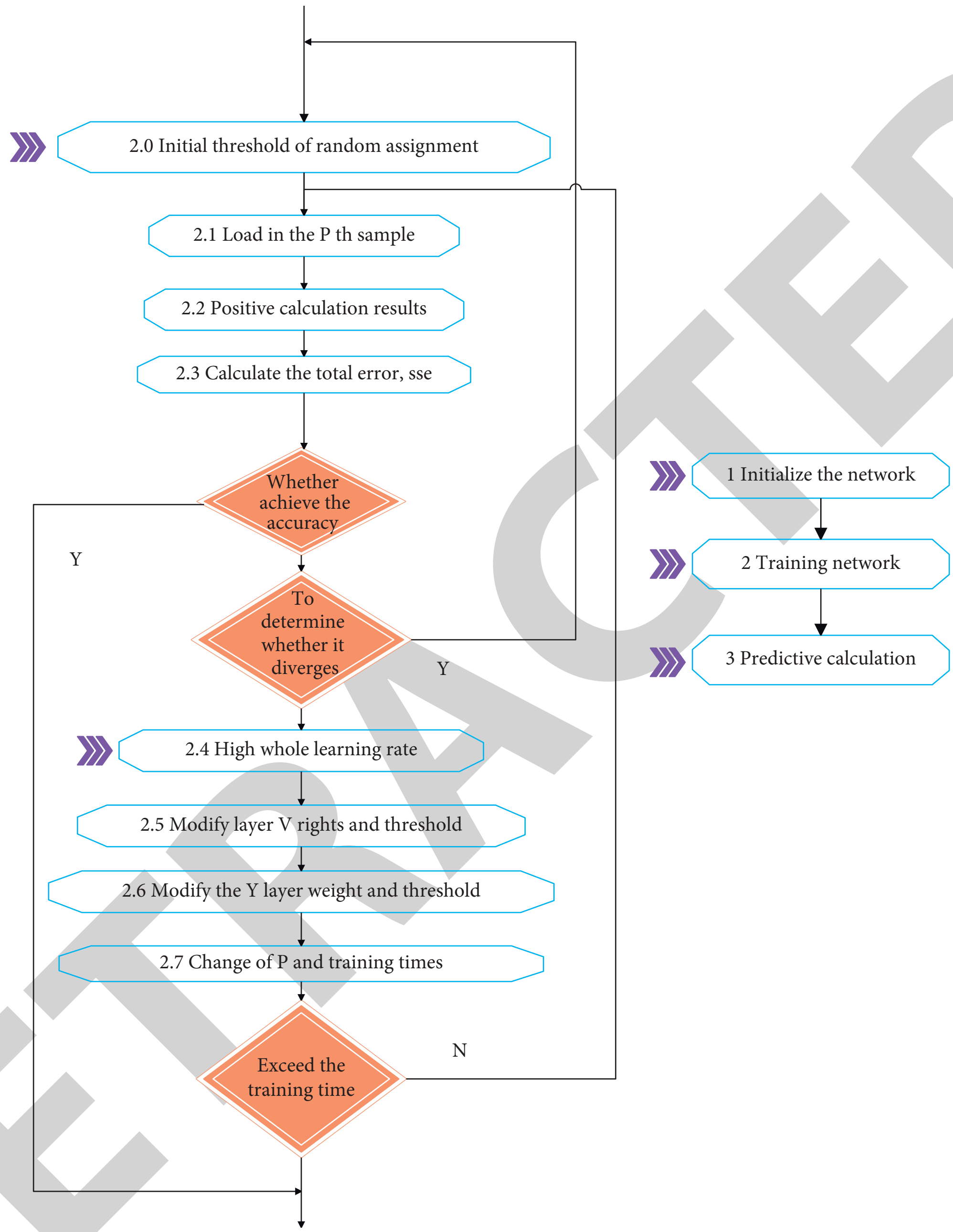
We build the following q-order SVAR model for each country in the sample:

$$\begin{pmatrix} 1 & a_{12} \\ a_{21} & 1 \end{pmatrix}\begin{pmatrix} PR_{i,t} \\ LR_{i,t} \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}\begin{pmatrix} PR_{i,t-1} \\ LR_{i,t-1} \end{pmatrix} + \Lambda + \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}\begin{pmatrix} PR_{i,t-q} \\ LR_{i,t-q} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i,1t} \\ \varepsilon_{i,2t} \end{pmatrix}. \quad (11)$$



(a)

Figure 2: Continued.



(b)

Figure 2: BP neural network model structure and training process. (a) Topological diagram of BP neural network. (b) Flowchart of the BP network training algorithm program.

Among them, $LR_{i,t}$ and $PR_{i,t}$ are the bank loan interest rate and the monetary policy interest rate of the sample country i, respectively. When y represents $(PR, LR)^T$, the above formula can be written more concisely as

$$A y_{i,t} = \Gamma_1 y_{i,t-1} + \Lambda + \Gamma_p y_{i,t-p} + \varepsilon_{i,t}. \tag{12}$$

If the matrix A is assumed to be nondegenerate, the corresponding simplified VAR can be obtained by multiplying both sides of the equation by A^{-1}:

$$y_{i,t} = A^{-1}\Gamma_1 y_{i,t-1} + \Lambda + A^{-1}\Gamma_p y_{i,t-p} + A^{-1}\varepsilon_t. \tag{13}$$

The disturbance term $\mu_t = A^{-1}\varepsilon_t$ of the simplified disturbance term is a linear combination of the structural VAR disturbance term.

This model is the "AB model" of SVAR. It is assumed that $A\mu_t = Be_t, \mu_t$ is a simplified disturbance term, e_t is a standard orthogonal random disturbance term, its constituent elements are mutually orthogonal, and the covariance matrix is an identity matrix. The characteristic of the AB model is that the current relationship of each endogenous variable in the system can be clearly established, and the impact of the standard orthogonal random disturbance term on the system can be analyzed intuitively. To identify an A B

Table 1: Correlation between interbank money market interest rate and financial crisis.

Number	Correlation
1	22.20
2	28.46
3	30.87
4	26.70
5	24.35
6	23.42
7	27.88
8	23.32
9	27.55
10	28.40
11	22.73
12	24.13
13	29.41
14	25.30
15	22.68
16	22.38
17	26.14
18	25.60
19	27.16
20	27.06
21	30.78
22	26.75
23	21.01
24	24.41
25	22.35
26	25.07
27	24.56
28	27.47
29	24.63
30	22.02
31	28.08
32	28.94
33	28.21
34	26.93
35	24.56
36	23.19
37	21.36
38	29.70
39	23.08
40	24.67
41	30.98
42	28.07
43	26.43
44	26.48
45	28.41
46	29.62
47	22.37
48	23.37
49	24.03
50	21.23

model with M endogenous variables, at least $[2M^2 - M(M+1)/2]$ constraints need to be imposed on the elements in matrices A and B. In this paper, the "Cholesky constraint" is used to set the matrix A as a lower triangular matrix, the main diagonal elements are all 1, and the matrix B is set as a diagonal matrix. From an economic point of view, such a setting is also in line with the general understanding that the monetary policy interest rate has a current

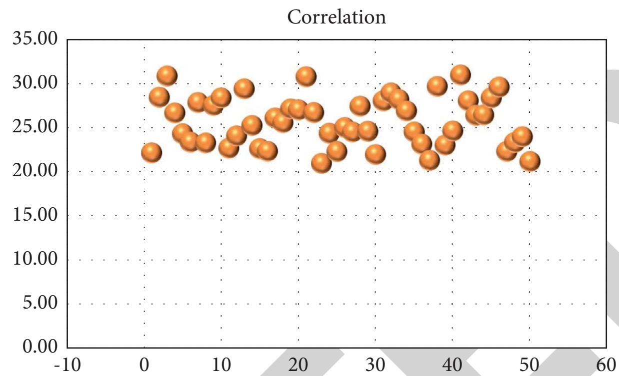


Figure 3: Statistical table of correlation between interbank money market interest rate and financial crisis.

impact on the loan interest rate, while the loan interest rate has no current impact on the monetary policy interest rate.

After the model is established, the dynamic response of the system under shock is analyzed by observing the impulse response function. This article treats giving a shock to the policy rate as the implementation of monetary policy, focusing on the impact of the shock on bank lending rates. According to the above SVAR model, the impulse response function of bank loan interest rate can be obtained when the shock to the policy interest rate occurs. The impulse response value of the first period (re_1), the average impulse response value of the first four periods (re_average), and the maximum impulse response value (re_max) after the shock are selected to reflect the response of bank loan interest rates to changes in monetary policy interest rates. By comparing the impulse response function values of different countries, we can see the difference in the transmission effect of monetary policy interest rate among different countries.

Next, the method of multiple regression is also used to study the influence of banking industry characteristics on the transmission effect of monetary policy interest rate, and the following multiple regression model is established. The explained variable is the impulse response value, and the annual average value of the corresponding sample period is also taken. In order to examine the impact of the financial crisis on the monetary policy transmission channels of various countries, this section also divides the sample into two parts before the financial crisis and after the financial crisis.

$$\begin{aligned}
\mathrm{re_1}_i &= \alpha_0 + \alpha_1 CR3_i + \alpha_2 \mathrm{npl}_i + \alpha_3 fd_i + \alpha_4 \mathrm{GDP}_i \\
&\quad + \alpha_5 \mathrm{CPI}_i + \alpha_6 rq_i + \xi_1, \\
\mathrm{re_average}_i &= \omega_0 + \omega_1 CR3_i + \omega_2 \mathrm{npl}_i + \omega_3 fd_i \\
&\quad + \omega_4 \mathrm{GDP}_i + \omega_5 \mathrm{CPI}_i + \omega_6 rq_i + \xi_2, \\
\mathrm{re_max}_i &= \lambda_0 + \lambda_1 CR3_i + \lambda_2 \mathrm{npl}_i + \lambda_3 fd_i \\
&\quad + \lambda_4 \mathrm{GDP}_i + \lambda_5 \mathrm{CPI}_i + \lambda_6 rq_i + \xi_3.
\end{aligned} \tag{14}$$

According to the method introduced above, we can obtain the impulse response function of the bank loan interest rate to the monetary policy interest rate in each sample

Table 2: Financial crisis prediction effect.

Number	Prediction effect
1	54.77
2	34.28
3	50.50
4	53.25
5	50.72
6	33.68
7	30.67
8	39.07
9	39.95
10	30.77
11	49.98
12	48.65
13	33.94
14	34.87
15	45.96
16	35.86
17	46.06
18	37.76
19	47.56
20	48.49
21	48.21
22	36.66
23	30.46
24	46.46
25	52.42
26	49.29
27	46.69
28	33.77
29	39.58
30	53.03
31	33.08
32	37.52
33	38.41
34	50.23
35	35.60
36	49.96
37	46.73
38	37.19
39	46.52
40	42.87
41	43.39
42	49.86
43	41.62
44	30.76
45	36.28
46	39.93
47	42.04
48	38.46
49	43.00
50	32.88

country before and after the financial crisis. The impulse response value re_1 of the first period after the shock, the average impulse response value re_average of the first four periods, and the maximum value of the impulse response function re_max are selected to reflect the monetary policy interest rate transmission effect of various countries before and after the financial crisis, as shown in Figure 1.

Figure 1(a)–1(f) shows the frequency distribution histogram of re_1, re_average, and re_max. The first-period

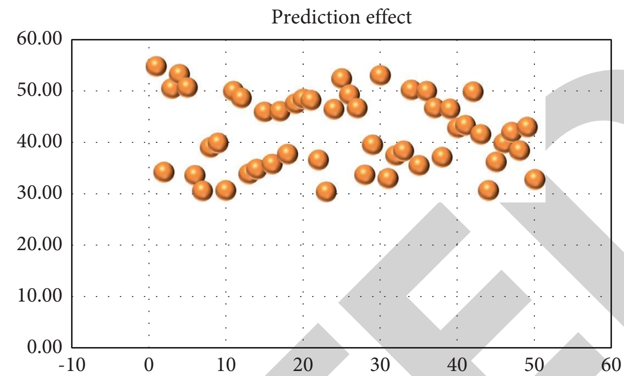


Figure 4: Statistical chart of the prediction effect of the financial crisis.

impulse response value (re_1) after a shock represents how the lending rate responds in the short term after a change in monetary policy. It can be seen from the figure that the re_1 indicators of most countries are positive numbers, indicating that changes in monetary policy interest rates will immediately cause the same direction changes in bank loan interest rates, which is consistent with the usual economic laws. However, comparing Figure 1(a) and 1(d), it can be seen that the re_1 indicator of countries generally decreased after the financial crisis, and the number of countries with this indicator less than 0 increased. Considering that there is a certain lag in the response of loan interest rates after the monetary policy interest rate changes in some countries, only using the impulse response value of the first period cannot fully explain the transmission effect of the policy interest rate to the bank loan interest rate. In this paper, the average impulse response value (re_average) and the maximum impulse response function (re_max) of the first four periods are also selected. Observing the frequency distribution histogram of these two indicators, it can be seen that these two indicators are positive numbers in most countries; that is, monetary policy has played a role in guiding bank loan interest rates to change in the same direction. However, after the financial crisis, the two indicators of re_average and re_max in various countries have generally decreased, which may mean that the transmission effect of monetary policy interest rates to bank lending rates has weakened after the financial crisis.

On the basis of the above analysis, this paper combines the neural network model to analyze the correlation between the interbank money market interest rate and the financial crisis. The BP neural network topology diagram and the BP network training algorithm program flowchart are shown in Figure 2.

In view of the shortcomings of traditional BP network, this paper improves it. First, the momentum term α ($0< \alpha <1$) is added when modifying the weights and thresholds, which plays an optimization role. Second, theoretically speaking, there can be infinitely many hidden layers in BP network, but generally only one layer is used in practical application. The improvement of network accuracy can be achieved by increasing the number of neurons. If we use the financial interest rate index of the previous n months to predict the financial interest rate index of the n+1st month, and so on, then only n ($n > 5$) node input layers and 1 node output layer are needed. As for the selection of the number of hidden layer nodes, the "trial and error method" is used in this BP network program, and it is found that the convergence is better when the number of nodes is 2n.

On the basis of the above research, the effect of the model proposed in this paper is verified. Moreover, this paper verifies the analysis effect of this model on the correlation between the interbank money market interest rate and the financial crisis through multiple sets of simulation data, and obtains the results shown in Table 1 and Figure 3.

From the above research, we can see that the model proposed in this paper can play a certain role in the correlation statistics between interbank money market interest rate and financial crisis. On this basis, this paper evaluates the prediction effect of the financial crisis and obtains the results shown in Table 2 and Figure 4.

From the above research, it can be seen that the correlation analysis between the interbank money market interest rate and the financial crisis based on the neural network model proposed in this paper can play a certain role.

4. Conclusion

From the analysis of the long-term and short-term effects, the marketization of interest rates will increase the risk of macroeconomic operation in the short term, but it will contribute to the increase of macroeconomic output in the long run. Interest rate liberalization may also deepen finance based on savings effect, income effect, channel effect, etc., which forces commercial banks to accelerate financial innovation and change their business methods. Moreover, the profits of commercial banks may shift to intermediary business, which will lead to substantial adjustment of the financial system and have a negative impact on the operation of the macro economy. However, most scholars believe that as long as the reform of interest rate liberalization does not cause a one-time large fluctuation in interest rates, the impact on the macro economy can still be controlled. This paper combines the neural network model to analyze the correlation between the interbank money market interest rate and the financial crisis. The experimental research shows that the correlation analysis between the interbank money market interest rate and the financial crisis based on the neural network model proposed in this paper can play a certain role.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This work was supported by 2016 Guangdong Distinctive Key Disciplines "public Administration" (F2017STSZD01).

References

[1] S. Barde, "Back to the future: Economic self-organisation and maximum entropy prediction," *Computational Economics*, vol. 45, no. 2, pp. 337–358, 2015.

[2] D. Bhattacharya, J. Mukhoti, and A. Konar, "Learning regularity in an economic time-series for structure prediction," *Applied Soft Computing*, vol. 76, no. 2, pp. 31–44, 2019.

[3] A. Ferramosca, D. Limon, and E. F. Camacho, "Economic MPC for a changing economic criterion for linear systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 10, pp. 2657–2667, 2014.

[4] Y. Geng, Z. Wei, H. Zhang, and M. Maimaituerxun, "Analysis and prediction of the coupling coordination relationship between tourism and air environment: Yangtze river economic zone in China as example," *Discrete Dynamics in Nature and Society*, vol. 2020, no. 10, 15 pages, Article ID 1406978, 2020.

[5] P. Karanikić, I. Mladenović, S. Sokolov-Mladenović, and M. Alizamir, "Retraction Note: Prediction of economic growth by extreme learning approach based on science and technology transfer," *Quality and Quantity*, vol. 53, no. 2, pp. 1095-1096, 2019.

[6] K. Ataka, "Prediction of election result and economic indicator," *Resuscitation*, vol. 96, no. 6, p. 84, 2014.

[7] S. Nagy and J Pipek, "An economic prediction of the finer resolution level wavelet coefficients in electronic structure calculations," *Phys.chem.chem.phys*, vol. 17, no. 47, pp. 31558–31565, 2015.

[8] J. Pipek and S. Nagy, "An economic prediction of refinement coefficients in wavelet-based adaptive methods for electron structure calculations," *Journal of Computational Chemistry*, vol. 34, no. 6, pp. 460–465, 2013.

[9] H. L. Vu, K. T. W. Ng, and D. Bolingbroke, "Time-lagged effects of weekly climatic and socio-economic factors on ANN municipal yard waste prediction models," *Waste Management*, vol. 84, no. 2, pp. 129–140, 2019.

[10] W. Yu and W. Huafeng, "Quantitative analysis of regional economic indicators prediction based on grey relevance degree and fuzzy mathematical model," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 2, pp. 1–14, 2019.

[11] L. Zhou, K. K. Lai, and J. Yen, "Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation," *International Journal of Systems Science*, vol. 45, no. 3, pp. 241–253, 2014.

[12] C. Teljeur, M. O'"Neill, L. Murphy et al., "Using prediction intervals from random-effects meta-analyses in an economic model," *International Journal of Technology Assessment in Health Care*, vol. 30, no. 1, pp. 44–49, 2014.

[13] P. Rajsic, A. Weersink, A. Navabi, and K. P. Pauls, "Economics of genomic selection: The role of prediction accuracy and relative genotyping costs," *Euphytica*, vol. 210, no. 2, pp. 1–18, 2016.

[14] F. Jahedpari, T. Rahwan, S. Hashemi et al., "Online prediction via continuous artificial prediction markets," *IEEE Intelligent Systems*, vol. 32, no. 1, pp. 61–68, 2017.

[15] V. Daksiya, H. T. Su, Y. H. Chang, and E. Y. M. Lo, "Incorporating socio-economic effects and uncertain rainfall in

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9862189, 1 page
https://doi.org/10.1155/2023/9862189

Retraction

Retracted: Application of Convolutional Neural Network Algorithm under Deep Learning in Digital Clothing Design

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Xue and X. Xue, "Application of Convolutional Neural Network Algorithm under Deep Learning in Digital Clothing Design," *Security and Communication Networks*, vol. 2022, Article ID 4880555, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4880555, 12 pages
https://doi.org/10.1155/2022/4880555

Research Article

Application of Convolutional Neural Network Algorithm under Deep Learning in Digital Clothing Design

Xiangyi Xue [1] and Xiangying Xue [2]

[1] *Apparel & Art Design College, Xi'an Polytechnic University, Xi'an, Shaanxi 710048, China*
[2] *University of Shanghai for Science a Technoilogy Foreign Language College, Shanghai 201109, China*

Correspondence should be addressed to Xiangyi Xue; 20100707@xpu.edu.cn and Xiangying Xue; xuexiangying@usst.edu.cn

Received 9 March 2022; Revised 2 April 2022; Accepted 15 April 2022; Published 24 May 2022

Academic Editor: Chin-Ling Chen

In order to overcome the influence of background, lighting, deformation, and other factors, using a constitutional neural network structure combined with metric learning, specifically, it includes two model structures, Siamese, and Triplet. The use of bicubic NURBS surfaces is proposed, the idea of constructing mannequins and garment pieces, the experimental results show that NURBS surface control is flexible and simple, and the calculation is stable, and it is the best surface for constructing virtual samples and avatars. Based on studying the three-dimensional structure design of clothing, based on the 10 key curves of the human body, the curve and surface interpolation algorithm is applied, by calling OpenGL related functions, the establishment of the benchmark human body model is well realized, and it lays a foundation for the deformation of the human body model based on parameters in the future.

1. Introduction

As competition in the global market intensifies, businesses are under increasing pressure. It has become an important issue for the survival and development of enterprises that it is necessary in the shortest possible time, with the lowest possible cost, to produce the highest possible quality product. The virtual product development technology is produced in such an era background [1]. It uses computers to complete product development, Based on life cycle modeling of products combined with computer graphics, artificial intelligence, concurrent engineering, network technology, multimedia technology and virtual reality, and other technologies that are integrated, in virtual conditions, conceive, design, manufacture, test, and evaluate products. One of its salient features is the use of digital models stored inside the computer—virtual products are used instead of physical models for simulation and analysis, thereby improving the decision-making level of products in multiple objectives such as time, quality, cost, service, and environment, and it forms a good rapid response mechanism with the market to improve the profitability of the product, to achieve the purpose of global optimization and one-time development success. The emergence of virtual product development technology, had a huge impact on the manufacturing industry. A more successful example is the application in the manufacture of Boeing 777 and European Airbus, its development cycle, from the original required 8 years and 4 years, respectively, down to 5 and 2.5 years today. Therefore, virtual product development technology is more and more widely valued by enterprises, and it has become a research hotspot in the current manufacturing industry. China's garment industry is a traditional industry with a low degree of modernization; most clothing design is still based on handmade design, compared with advanced garment companies in other countries, and there is a large gap in design capability and brand effect. With the reform and opening and the development of computer technology, clothing enterprises have also introduced advanced technology and high-tech products one after another, especially the application of garment CAD system in the garment industry, the design and production of garment enterprises have brought great economic and social benefits [2]. However, fashion designers have always hoped that the display of three-dimensional effects can be realized on the computer, as shown in Figure 1. At present, whether manual design or computer design, none of them have solved the key technology of 2D template to 3D effect display. Apparel design, including style design, sample design, and the dressing effect can only be observed after the actual sample clothes are actually sewn, and there is a certain blindness and lag. If the expected effect is not achieved, it can only be modified and sewed many times, which is time-consuming, laborious, and expensive. Therefore, if "virtual sample clothing generation" can be realized, that is, "conversion and display of 2D template to 3D effect," it can not only solve the above problems well but also enable enterprises to adapt to the development trend of the modern clothing market, that is, multiproducts, small batches, fashion, and personalization, providing technical support for rapid response. At present, the two-dimensional garment CAD technology is relatively mature. Based on 2D, CAD technology has been applied to clothing design and production, which greatly shorten the development cycle of clothing products to a certain extent, and the contradiction between the long prediction cycle of clothing popularity and the short consumption cycle has been alleviated. But for the effect of the design, especially whether the model design (structural design) fits well, in the existing clothing CAD system, and whether it is a clothing style design system or a structural design system, none of them have addressed or solved this problem [3]. The display of the clothing effect of the clothing design belongs to the category of 3D clothing CAD technology. The 3D digitization technology of clothing is a hotspot of concern and research in the academic and business circles at home and abroad in recent years. In the 2002 National 863 Plan, in the field of advanced manufacturing and automation technology, a special topic of digital design and manufacturing has been opened, and it is required to focus on the development and industrialization of 3D digital design system, to carry out innovative research on several key technologies in digital design system and manufacturing. In addition to virtual clothing style design and virtual clothing structure design, virtual clothing design also includes virtual garment craft design. Optimal arrangement of garment production lines is an important content of virtual garment craft design. The important task of the clothing production line arrangement is to coordinate the human, material, financial, and energy aspects of the clothing sewing line. At present, all kinds of garment enterprises do not have a unified design model for the arrangement of production lines, usually only relying on the experience and feeling of the process planner, making manual estimates and piecing together, lack of knowledge, regularity, and randomness is relatively large, not only is the workload huge and time-consuming, and because there is no prior forecast, it increases the production cycle and production cost of the product, it restricts the improvement of economic benefits, which is incompatible with the constantly developing modern garment manufacturing industry. If the clothing production line can be optimized and arranged, it not only frees process planners from heavy and repetitive manual labor but also shortens the design cycle, guarantees the quality of design, improves the standardization of process design of garment enterprises, improves the production efficiency, technical level, management level, and economic benefits of the garment industry [4].

2. Literature Review

Digital clothing technology is the product of a combination of digital technology and traditional clothing design and production, and the use of digital technology, the design, production, marketing, delivery, and other links of clothing enterprises have been transformed and improved [5]. Puzyrev et al. found that 3D anthropometry is the basis of human and clothing modeling, 3D anthropometric techniques fall into two categories: One is the laser-based scanning technology (laser-based), and the other is the moire-based projection technology (moire-based) [6]. Based on the technical principle of laser, the three-dimensional image is obtained by using the laser scanning triangulation method. Ripple-based technology uses a white light source to project sinusoidal fringes onto the surface of an object, through the target object, and the phase transition of the projected light occurs to describe the surface contour of the object. The representative product of the former is the American Cyber-ware system, and the latter is the PMP system of the American TC2 company. How to extract the feature size required for clothing design from 3D scanning data is an area to be improved [7], Anton et al. developed a software to extract dimensions from digitized images, and it has been applied in the US military [8]. Shan et al. used feature-based technology to extract the feature size required for clothing from 3D human model data. A 3D human body modeling and clothing has always been a hot and difficult point in the field of computer graphics and CAD clothing [9]. For a long time, research in this field mainly formed the following modeling methods: use points, lines, and curves to build a 3D wireframe model, and use voxels to build a 3D solid model; use points, edges, and surfaces to build a 3D surface model, using the mesh facet method; and the 3D surface model is built on the basis of 3D physical modeling. Albelwi et al. found that there are additional neural networks for the exploration of freeform surface reconstruction. These methods have their own advantages in terms of operation speed, model controllability, and model smoothness, for example: The wireframe model has a fast calculation speed, but the model has a poor sense of reality, prone to ambiguity. The solid model is better in modeling efficiency and realism; however, the fitting of the human body surface shape is not ideal, and the topological relationship between voxels is complex: The mesh facet method is simple and effective, and with the increase of the number of facets, the simulation effect of the surface is better [10]. Ni et al. found that due to the large number of data points, the operation speed will be slowed down, and it is difficult to locate and control a single data point. The surface method uses parametric surface equations to build models. The data control points are few, the operation speed is fast, and the surface can be continuous in any order, and the information

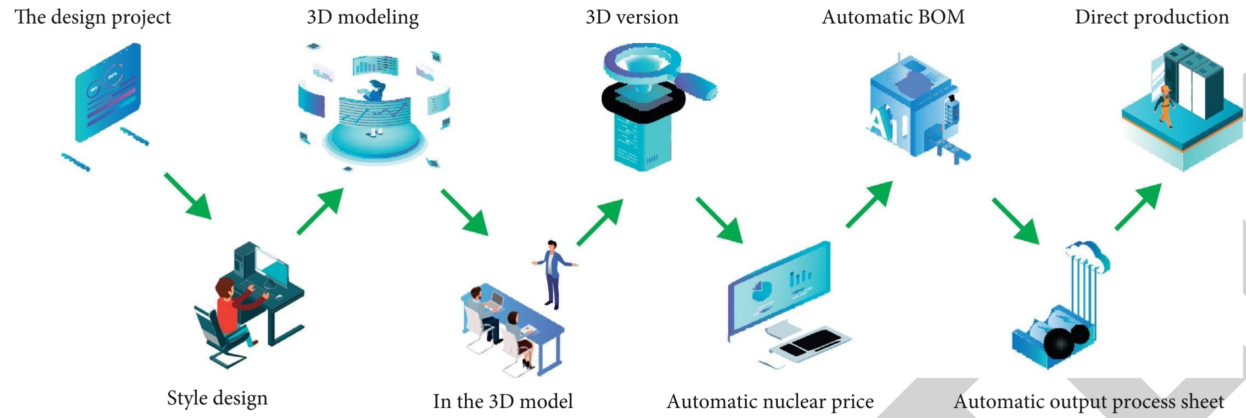


Figure 1: Flowchart of digital clothing design.

of the points on the surface is easy to obtain. In physically based 3D modeling methods, the physical information of the human body and the external environmental factors (such as gravity) where the human body is located are introduced into traditional geometric modeling methods; therefore, a more realistic modeling effect can be obtained; however, compared with traditional geometric modeling methods, it is much more complicated in operation. Neural network surface reconstruction has the advantages of high fitting accuracy to the measurement data, fewer surface patches, and editing and modification of local surfaces, but the problem is how to choose network parameters reasonably, solving the contradiction between network training speed and approximation accuracy, and solving problems such as continuous splicing which deserves further study. Therefore, in computer 3D clothing design including 2D and 3D mutual conversion, there is a need to do a study of fabric texture mapping, analysis of optical and mechanical properties, and 3D interactive design [11]. Bangari et al. found that many institutions at home and abroad are currently engaged in the research and development of 3D CAD apparel. From the perspective of 3D garment CAD system and application, PAD system has a certain conversion function from 2D to 3D, but its generated 3D clothing lacks realism: Lectra's high-end CDI-U4Ia already contains three-dimensional technology, and partially realizes the function of transforming 3D designs into 2D pieces, allowing designers to cut designs; however, its technology is not yet perfect, first of all, it has extremely high requirements on hardware configuration, and requires a workstation configuration, and ordinary PCs cannot be used at all. Second, its simulation effect needs to be further improved [12]. In addition, Fashion Studio System of Dynamic Graphics of Canada can produce more realistic three-dimensional clothing than the Maya Clothing module of Alias/Front Company in the United States; it also has powerful modeling functions of flexible bodies such as clothing and fabrics, but these systems are geared toward clothing animation, not clothing production of 3D clothing systems, and there is a certain distance from practical application. In the 3D clothing generation technology, an important link is the generation and stitching of garment pieces. In this process, 2D clothing cutouts are sewn onto a 3D mannequin, and form the initial shape of the three-dimensional clothing. Luo et al. adopted the elastic deformation model in which the clothing surface is discretized into a particle system and by solving the differential equation of the space motion of the particle system, we get the evolution of the system from a time series. Its research focuses on the dynamic simulation of fabrics, introduces external constraints, and controls the stitching of 2D garment pieces to 3D garments [13]. Using the energy approach, Patel et al. map the two-dimensional clothing piece to the three-dimensional mannequin, forming a joined rigid surface, and the mechanics of the fabric are characterized as energy equations. The method is constrained by the human model, and the large deformation prediction is carried out with the minimum energy at each point in space, gets the shape of the 3D garment in equilibrium, and it is suitable for expressing the static effect of 3D clothing [14]. Jagtap et al. studied virtual humanoid clothing, both of which adopted the classical proton-spring model. The description of the mechanical properties of the fabric is simple and clear; however, the fabric is required to be meshed on four sides according to the warp and weft directions, which bring certain difficulties to the sewing of complex garments [15]. Lin et al. proposed a two-dimensional and three-dimensional mapping algorithm based on the spring mass deformation model and considered the problem of interference checking [16].

3. Methods

Clothing virtual design facilitates the fashion designers to put the ideas in their minds, and it can be realized accurately and quickly in the computer through virtual technology, and the three-dimensional effect can be seen without the need to make samples. Therefore, it can not only shorten the production cycle and saves production costs but also enables products to enter the market as soon as possible, better highlights the fashion features of clothing, considers the protection of design patents, reduces business risk. Clothing design consists of three parts: clothing style design, clothing structure design, and clothing technology design; therefore, the virtual clothing design should include these three contents. With the proliferation of digital images under the rapid development of Internet information technology, image retrieval method based on text description and traditional method of extracting resources based on content have gradually been unable to meet people's needs; therefore, many scholars have turned their attention to deep learning, which has unique advantages in the field of images; using the deep learning convolutional neural network method, the image is introduced into the neural to conduct research related to image classification and retrieval [17].

Deep learning is one of the most popular branches of machine learning, covering many fields such as neural networks, graph modeling, pattern recognition, optimization theory, and signal processing. With a focus on business and education, the model of deep learning is a neural network device based on the study of the structure of the axonal communication network formed by the millions of neurons in the human brain and transforms the data in a similar way—simulation. Brain data is collected by simulating this structure, and the first layer is propagated to the next layer. Each layer works differently. Deep learning models differ from shallow models such as support vector systems and Markov hidden models, as well as deep neural network models for automatic learning. Undertaking Developer Neural Networks solves the problems that can change the neural networks mapping. Neural network refers to the potential of neural network to change. On the given sample set {x, y} according to the neuron principle of deep learning, the mapping transformation structure of a single neuron to the input and output is shown in Figure 2:

A single neuron model is also called a logistic regression model:

(1) Input and output:

In the neuron model of Figure 2, the input is x and the output is the following:

$$f(x) = f\left(\sum_{i=1}^{5} w_i x_i + b\right). \quad (1)$$

(2) Activation function

In Figure 2, (f)(x) is the activation function, which generally chooses the Sigmoid function, the hyperbolic tangent Tanh function, and the expressions of Sigmoid functions such as ReLU function, Softmax function, and Tanh function that are shown in formulas (2) and (3), and the image is shown in Figure 3:

$$f_1(z) = \frac{1}{1+e^{-z}}, \quad (2)$$

$$f_2(z) = \tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}-e^{-z}}. \quad (3)$$

As can be seen from the image, the relationship between the Tanh function and the Sigmoid function is shown in the following formula:

$$\tanh(x) = 2\ \sin\mathrm{moid}(2x) - 1. \quad (4)$$

The derivative of the sigmoid function directly uses its own output value, which is convenient for derivation; however, because it outputs the result to [0, 1], and both ends tend to be saturated, resulting in the saturation of the

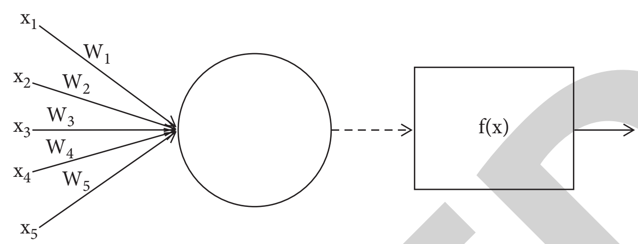


Figure 2: Structure of a single neuron model.

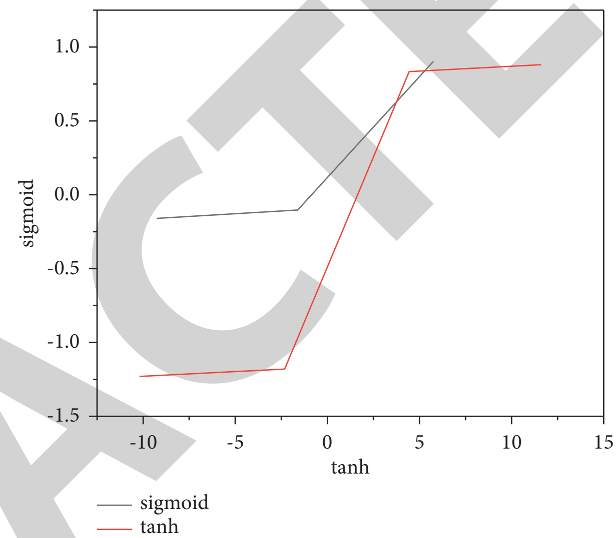


Figure 3: Image of Sigmoid function and Tanh function.

gradient, which is prone to the problem of gradient disappearance. Tanh function has faster convergence speed, and the output is centered at the origin, but there is still the problem of gradient disappearance, and the ReLU function effectively solves the problem of gradient disappearance. Currently, ReLU is generally selected as the activation function. Extend the single neuron model, backpropagation with chain derivation and gradient descent, and it constitutes a general artificial neural network. CNN convolution neural networks are special, deep feedforward artificial neural networks with translations, scaling, and tilt invariance, and a neural network formed by nonperfect connections between multiple networks. The overall structure of the convolutional neural network is shown in Figure 4:

In the overall structure of the convolutional neural network, it is mainly composed of a input layer, an output layer, and a hidden layer; the hidden layer only uses the convolution kernel, as the base convolution operation in the form of filters and accepts the full connection later. Convolution layer is used to reduce the loss of data feature, so it is called convolutional neural network. Convolution type neural networks have excellent results in image fields because of convolution of the design structure, weight sharing, and pooling operation to pile convolution and pooling layers, and convolutional neural networks have characteristics of image features and translation invariance. Image transfer between layers to extract functions can save enough image attributes. The core of the convolution type neural network is extracting the characteristics of the

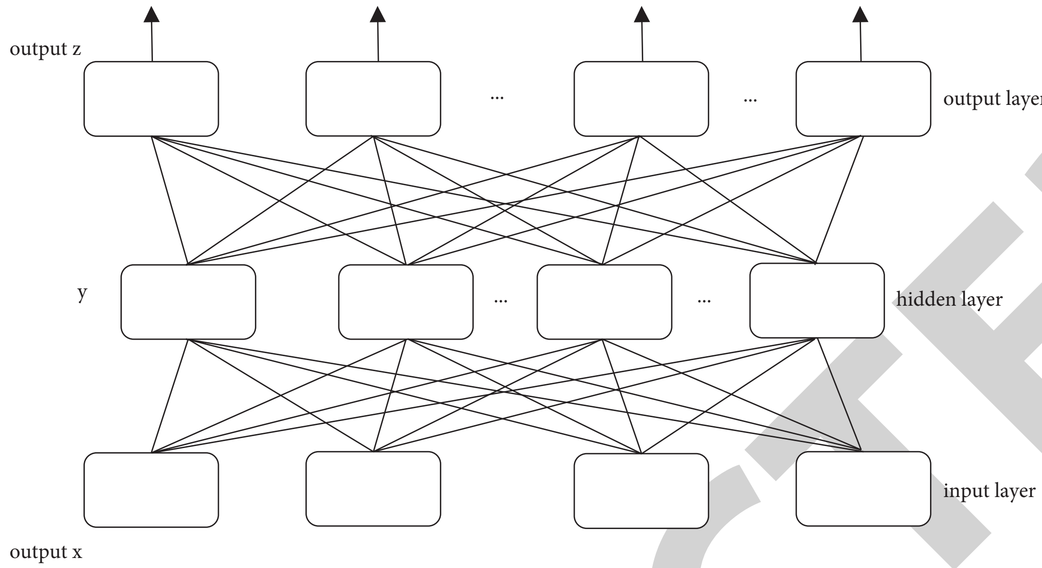


Figure 4: Overall architecture of convolutional neural network.

convolution level and pooling level, and the convolution level interacts with the input data to extract the characteristic through the convolution kernel, and the pooling level weakens location information and high-frequency important information is removed. Generates more abstract features and extracts layers and combines layers from full description layer, weight sharing, and pooling features to reduce the number of parameters. Convolution neural network with excellent feature extraction performance is required to increase the learning speed. Convolution neural networks are generally composed of data input layers and convolutional layers. ReLU has excitation layer, pooling layer, fully connected layer, and output layer [18]. Convolutional neural networks are mainly composed of some convolutional layers with a certain number of channels stacked, and the convolution operation of incoming data is checked by convolution to perform feature extraction and feature mapping and weighted summation of the incoming activation layers with multiple feature maps, to get the feature map for nonlinear feature extraction. The formula of the convolution operation is shown in equation (5):

$$s(i, j) = (X * W)(i, j) = \sum_m \sum_n X(m, n) W(i - m, j - n). \quad (5)$$

Convolutional neural networks generally use SAME convolution operations and VALD convolution operations for image data of input size N $*$ N; the convolution kernel is F $*$ F, if the step size is S, and the output size of the SAME convolution operation, and the VALD convolution operation is shown in formulas (6) and (7):

$$\text{Height} = \text{Weitht} = \frac{N}{S}, \quad (6)$$

$$\text{Height} = \text{Weight} = \frac{N - F + 1}{S}. \quad (7)$$

The VALD convolution operation of the convolutional layer is shown in Figure 5:

The excitation function of CNN generally adopts rectification linear unit (Relu), and Relu is easy to compute, have fast convergence velocities, easy to find gradients related to Sigmoid function and Tanh, and the problem of gradient disappearance is effectively relaxed and may form deeper networks [19]. The relay is used to correct the linear unit as an excitation function to perform B4 nonlinear mapping on the convolution layer output results. As shown in formula (8)

$$f(x) = \max(0, x). \quad (8)$$

Figure 6 shows the image of the Relu activation function:

In general, for convolutional layers with small receptive fields and strides, the feature maps obtained after convolution are larger. This time places a pooling layer in successive convolutional layers, and on the premise of ensuring that the depth features remain unchanged, sampling is performed. This method is used to reduce the dimensionality of map features. The characteristics data generated by the convolution level are sampled for dimensional reduction and compression of regional characteristics. Figure 7 shows the pooling operation of the pool layer.

Typically, after the last grouping layer, a plurality of fully connected layers is connected as hidden layers [20].

The key to learning lies in setting up the training process, and the weights between the layers of the neural network; the mapping relationship is updated by updating the weights. It is used in the forward propagation process to calculate the partial derivative of the loss function for each weight in the gradient derivative, and the gradient is updated according to the gradient descent equation to optimize the net to optimize the back propagation. In forward propagation, if the weight of each layer of the net is w for the sample passed to the neural net, the mapping function is f, and after passing through each layer, the output result is the matrix multiplication of the sum weight, reaching the output layer. The output equation is shown in formula (9):

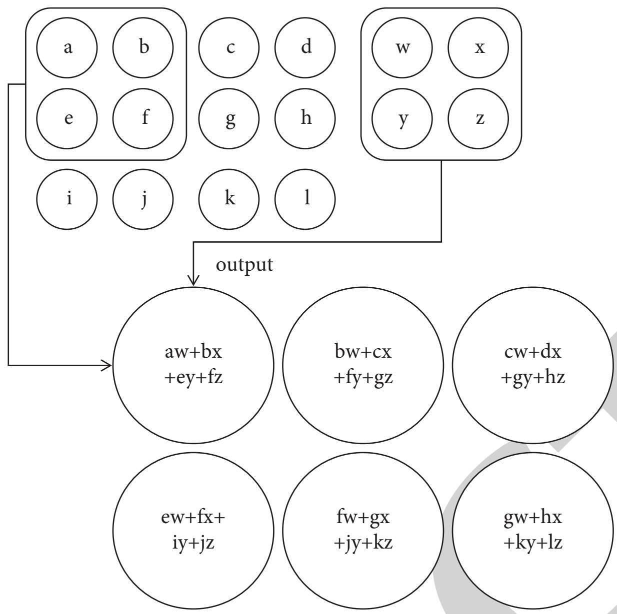


Figure 5: VALD convolution operation.

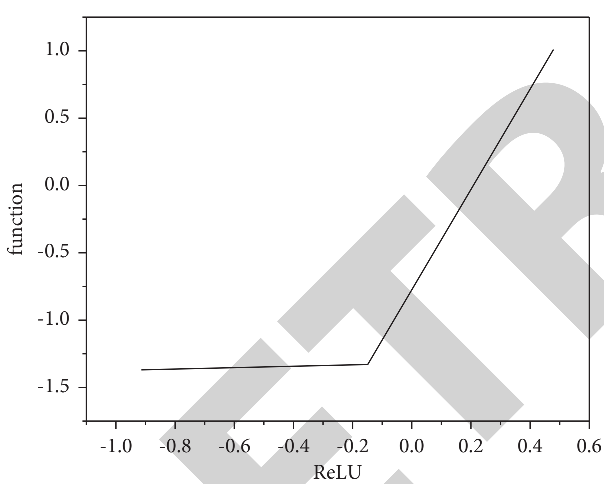


Figure 6: Relu activation function image.

$$o_p = F_n\left(\cdots\left(F_2\left(F_1\left(X_p W^{(1)}\right)W^{(2)}\right)\cdots\right)W^{(n)}\right). \tag{9}$$

Backpropagation refers to the rate of output from the neural network, which is propagated by the error of the actual output rate, adjusting the weight coefficient obtained in the forward propagation process, and reducing operating costs. Sample backpropagation error formula is as follows:

$$E^n = \frac{1}{2}\sum_{k-1}^{c}\left(t_k^n - y_k^n\right)^2 = \frac{1}{2}\left\|t^n - y^n\right\|_2^2, \tag{10}$$

where C is the total number of categories in the sample set. t_k^n K-th classification cup in the sample data, y_k^n. The k-th output data corresponding to the k-th sample value.

The update process of the weights by the backpropagation algorithm is as follows:

For a given sample set $D = \{(x, t)\}$, initialize network structure $d \times n_H \times c$.

Repeat the loop multiple times until the expected average misunion is achieved for the sample set. Calculating the error :

$$J = J_x(\omega) = \frac{1}{2}\sum_{K-1}^{C}\left(t_k - z_k\right)^2. \tag{11}$$

GoogleNet is the 2014 ILSVRC classification challenge and is the CNN model that won the championship with a Top5 error rate of 6.67%. It has a faster convergence speed than VGGNet, effectively shortening the training time, and there are 22 layers of network, including 21 convolutional layers and 1 fully connected layer. Local densification of sparse matrix operations in the form of multiscale convolution, after the features are processed in parallel, and feature concatenation is performed [21]. The 1×1 convolution kernel is used for dimensionality reduction, which effectively improves the computational efficiency. The model component of the Inception Module and the different scales of perceptual fields are obtained through different convolution kernels, then, data splicing improves the feature expression ability of the network, approximate to a CNN with a dense structure. The Inception Module structure is shown in Figure 8:

For the performance evaluation criteria of image classification and retrieval, the standard for evaluating the quality of classification performance generally considers the classification accuracy, the evaluation of retrieval performance mainly includes retrieval accuracy, sorting effect, and retrieval speed. The accuracy rate reflects the performance of the feature extraction algorithm and the similarity matching algorithm, and the sorting effect and retrieval speed reflect the image feature indexing effect and the complexity of the similarity matching algorithm.

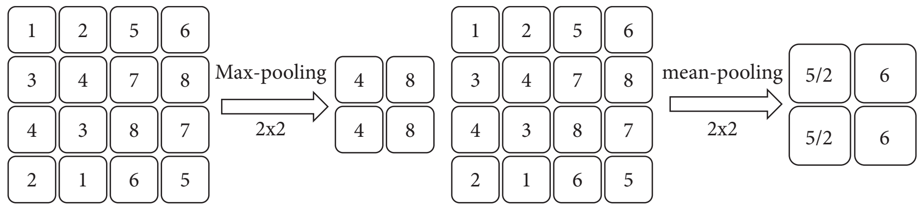


FIGURE 7: Max pooling (left) and mean pooling (right) process.

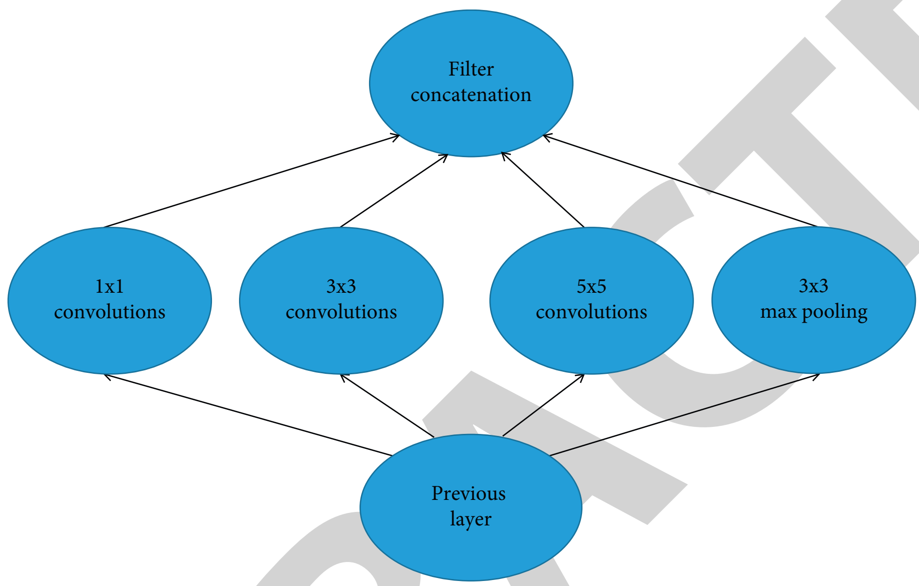


FIGURE 8: Inception Module structure.

For classification and retrieval accuracy, generally, precision, recall, and precision are used to evaluate. Precision is a measure of retrieval accuracy, which is defined as follows: during a retrieval process, the system that returns the search results. The total number of images is returned as the percentage of the correct images. Recall is used to represent the ratio of the number of correct images to the number of correct images in the image library to get the information defined as a search result. The precision points to the percentage of the total number of returned correctly. General consideration parameters are shown in Table 1:

The formulas for precision, recall and precision are as follows:

$$\begin{aligned}
\text{Precision} &= \frac{\text{Number of correct images in retrieved images}}{\text{The number of correct results in the system}} = \frac{\text{TP}}{\text{TP} + \text{FP}},\\
\text{Recall} &= \frac{\text{The number of correct results retrieved}}{\text{The number of all correct results in the library}} = \frac{\text{TP}}{\text{TP} + \text{FN}},\\
\text{Accuracy} &= \frac{\text{Number of correct images in retrieved images}}{\text{Number of all retrieved results}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.
\end{aligned} \tag{12}$$

Usually, the ranking evaluation method and the matching percentage are used as a measure of the sorting performance of retrieval results.

The recall rate and precision rate do not consider the position of the returned relevant image in the retrieved image, and it is impossible to judge the quality of the search result sequence. The average sequence number of the retrieved related images can be used to solve this problem, for TP, FP, FN, TN in Table 1, o is the sequence number of the retrieved related pictures in the retrieved pictures, let K represent the average ranking number of the relevant images returned by the system when retrieving:

$$K_1 = \frac{1}{\text{TP}} \sum_{r=1}^{\text{TP}} O_r. \tag{13}$$

Table 1: Performance evaluation criteria.

	Relevant	NonRelevant
Retrieved	TP	FP
Not retrieved	FN	TN

For the best algorithm implementation, the more relevant the image in the retrieval result is, the higher the sequence number is, that is, the ideal average sequence number is the following:

$$K_2 = \frac{TP}{2}. \tag{14}$$

The retrieval performance of a single relevant image was evaluated using the matching percentage, and for ideal retrieval algorithm results, the target image should be ranked first; however, it is difficult to achieve this retrieval effect in the actual situation to evaluate the performance of the system retrieval for a single image, which is defined as follows:

$$M = \frac{N - S}{N - 1} \times 100\%, \tag{15}$$

where N is the number of results returned by the system during retrieval, S is the order of the target image in the returned retrieval results, and usually, the average value of multiple experiments is used to indicate the quality of retrieval effect.

The model needs to perform feature extraction, similarity measurement, and sorting of clothing images during retrieval, using different feature extraction algorithms and similarity matching algorithms, searches that return the same number of search results will result in different search times; therefore, the retrieval speed is defined as the ratio of the number of retrieved results to the retrieval time. For the data in Table 1, in the retrieval process of a batch, if the retrieval time is T, the retrieval speed is defined as follows:

$$\begin{aligned} \text{retrieval speed} &= \frac{\text{Number of results to retrieve output}}{\text{Search time}}, \\ &= \frac{TP + FP}{T}. \end{aligned} \tag{16}$$

4. Results and Discussion

The results of the process analysis can be represented by a process flow diagram. In the industrialized production of garments, the basic data of the assembly line production arrangement are the process flow chart. The process flow chart is based on the before-and-after relationship of garment processing, and the processing sequence and time are expressed graphically [22]. The process flow chart should include the name of the processing process, the processing time (pure processing time or standard processing time), and the type and model of the equipment used (as shown in Figure 9). Usually, various fixed graphic symbols are used to distinguish the operation nature of each process (Figure 10).

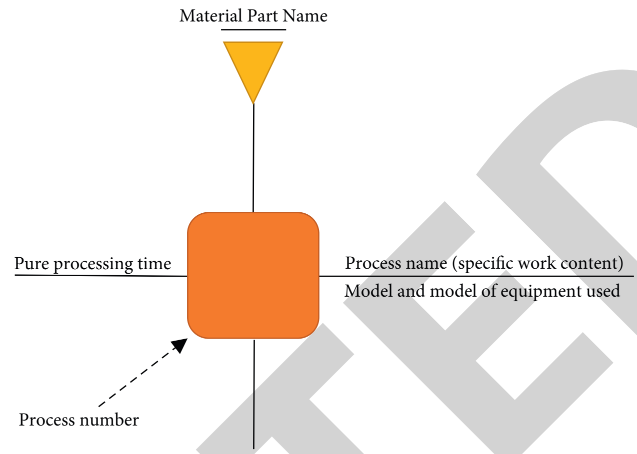


Figure 9: Process flow chart expression.

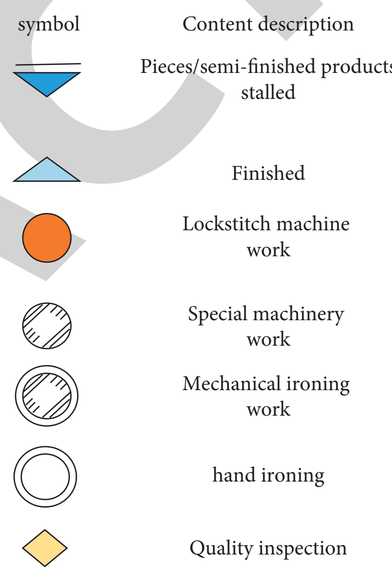


Figure 10: Graphical symbols of each process.

According to actual needs, enterprises can also formulate certain symbols by themselves.

The general flow of the neural network algorithm is shown in Figure 11:

Step 1. Randomly generate the initial population, the number of individuals is certain, and everyone is represented as the gene code of the chromosome.

Step 2. If it matches, you will meet the optimal criteria, calculate individual fitness, and the judge will output the best individual and its representative optimal solution and exit the calculation, otherwise to the third step.

Step 3. Choosing individuals updated according to fitness, high fitness individuals can have selected high probability and eliminate low fitness individuals.

Step 4. Create new individuals according to certain crossover probability and crossover methods.

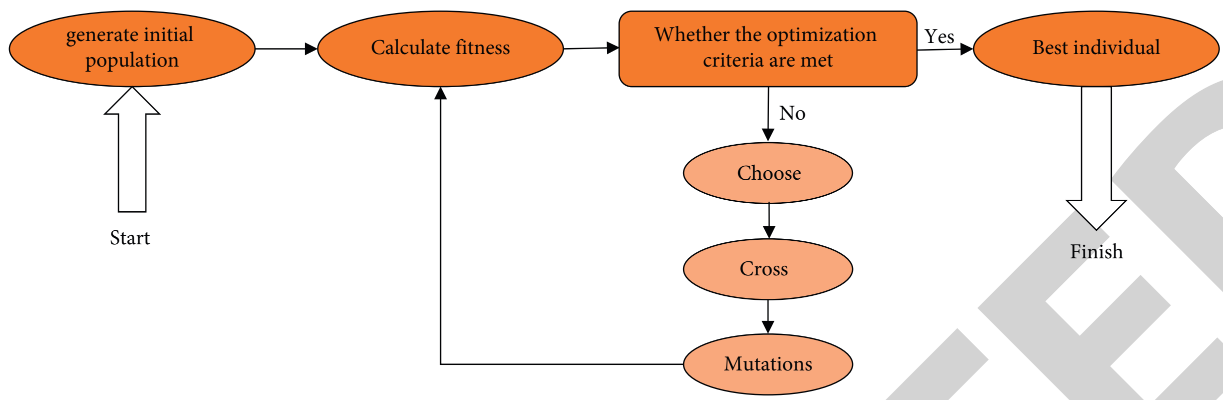


Figure 11: Flowchart of the convolutional neural network algorithm.

Step 5. Generating new individuals according to certain mutation probabilities and mutation methods.

Step 6. Generate a new generation of population by crossover and mutation, and return to Step 2.

In the process of recombination, the randomly generated chromosomes are rearranged, making it conform to the sequence of processing procedures, and in order to facilitate the calculation of the fitness function, generate an array of job numbers and machines. Among them, the array job number records the branch number of each process and its relative position in the branch, the array jiqi records the processing equipment number and tool number of each process. The main flowchart of the reorganization operation is shown in Figure 12:

When running the quadratic selection convolutional neural network algorithm program, some parameters need to be selected in advance; they include population size, crossover rate, normal mutation rate, inbreeding mutation rate, evolutionary generation, etc. These parameters have an important impact on the performance of the quadratic selection convolutional neural network algorithm. For specific problems, whether the measurement parameters are set appropriately or not, it should be judged based on the convergence of multiple runs and the quality of the solution [23]. If it is difficult to adjust the parameters to effectively improve the performance of the convolutional neural network algorithm, and it is often necessary to improve the convolutional neural network algorithm again. Taking the process flow diagram of Figure 13 as an example below, the optimized process arrangement scheme is given, and the parameters set are shown in Table 2:

Figure 14 shows individual fitness (maximum and average) evolution curves over generations.

From Figure 13, it can be seen intuitively that from the optimized process arrangement plan, the processing time of the whole process as 405 s is obtained, and the waiting time of the equipment during the processing is 0s. In Figure 14, the abscissa represents the number of iterations, the ordinate represents the fitness value, and the upper curve represents the best fitness in each generation of the population, and the curve below represents the average fitness of the population across generations. It can be seen that the algorithm produces a local convergence phenomenon before the 10th

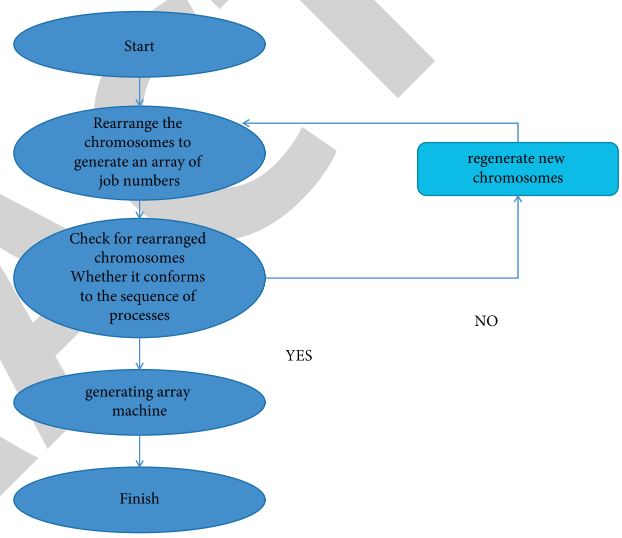


Figure 12: Main flow chart of reorganization operation.

generation, but after several generations of evolution, it jumped out of this local convergence zone, and the superior performance of the quadratic selection convolutional neural network algorithm was confirmed [24].

The algorithm adopts the quadratic selection strategy and the expanding population sorting method, and the fitness curve of the past generations of the quadratic selection convolutional neural network algorithm is shown in Figure 15, and the parameter settings of the algorithm are shown in Table 3. In this algorithm, when the population size is 200, the number of selected crossover individuals is only 100, and the number of mutant individuals is only 50, the remaining 50 individuals are directly added to the matching set. Therefore, when the crossover rate is 1, the number of crossover parent individuals is only 100, and the corresponding crossover rate of the standard convolutional neural network algorithm is $100/200 = 0.5$. When the mutation rate is 0.5, the number of mutated parental individuals is only about $50 \times 0.5 = 25$, and the corresponding standard convolutional neural network algorithm has a mutation rate of $25/200 = 0.125$.

As can be seen from Figure 15, after running for 100 generations, the fitness of the best individual is 18, and the

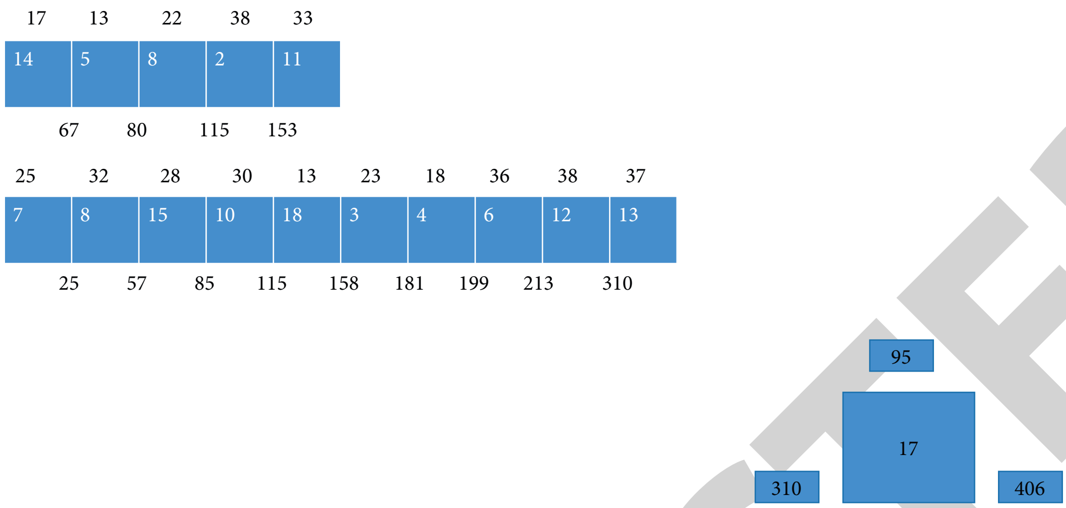


Figure 13: Shows the optimization scheme of the process choreography represented by the Gantt chart.

Table 2: Parameters of the quadratic selection convolutional neural network algorithm.

Population size (popsize)	100
Evolutionary algebra (generation)	50
Crossover rate (pc)	1
Normal variation rate (Pm)	0.5
Consanguineous variation rate (pmr)	1

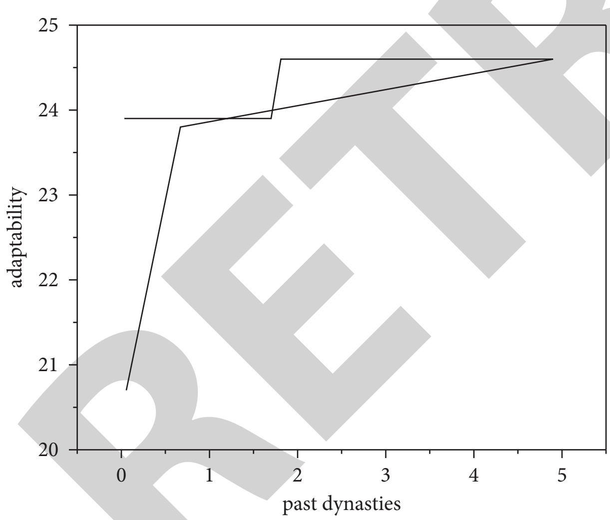


Figure 14: Fitness curve of past generations.

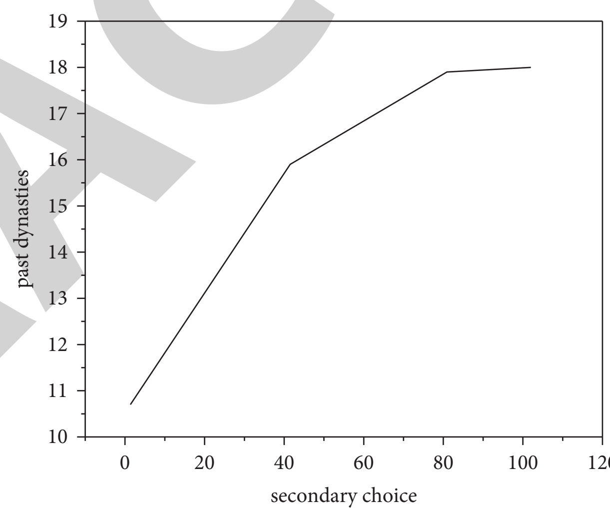


Figure 15: The fitness curve of successive generations of the quadratic selection convolutional neural network algorithm.

Table 3: Quadratic selection convolutional neural network algorithm parameters.

Population size (popsize)	200
Evolutionary algebra	100
Crossover rate	1
Normal variation rate (Pm)	0.5
Consanguineous variation rate (pmr)	1

average fitness is also 18, compared with the previous algorithm, the fitness and average fitness of the best individual obtained by this algorithm are the highest. The convergence of the algorithm is good, and after running the algorithm several times, the result converges to the global optimal solution. In addition, when local convergence occurs, although the individuals in the population are very close; however, since the algorithm adopts the mutation operation on close relatives, the new individuals produced by crossover have obvious changes, which are quite beneficial to the evolution of the population. And even if there is a local convergence phenomenon, the algorithm can also perform mutation operations, and the diversity of the population is maintained, so that the algorithm can jump out of the local convergence zone. From the above comparison, the quadratic selection convolutional neural network algorithm has better stability and convergence, and obtained satisfactory optimization performance [25].

5. Conclusion

Introducing relevant theories of deep learning, it focuses on the relevant principles and commonly used model components of convolution neural networks used by authors, including convolutional layers, pooling layers and activation functions, and the gradient calculation method of each component in the convolutional neural network. Expounds the concept, method, and content of virtual design and manufacturing, on this basis, and the main contents and models of virtual clothing design are analyzed and studied, a key issue in 3D garment CAD is researched emphatically, that is, the conversion of a 2D model to a 3D effect display. Based on the theory of computer graphics and clothing structure design, a method of converting 2D to 3D display based on the idea of decision is proposed, and an implementation plan is given. According to the characteristics of human body and clothing, the methods and characteristics of various curved surfaces are analyzed and compared, and the idea of using the bicubic NURBS surface to construct the human body model and the clothing piece is proposed. The experimental results show that the NURBS surface control is flexible and simple, the calculation is stable, and it is the best surface for constructing virtual samples and avatars. Based on studying the three-dimensional structure design of clothing, based on the 10 key curves of the human body, the curve and surface interpolation algorithm is applied, and OpenGL related functions are called, the establishment of the benchmark human body model is well realized, and the foundation for the deformation of the parameter-based human body model in the future is laid. Human body model building method, compared with the human body model construction method in the relevant literature, has the characteristics of good surface fitting effect, fast calculation speed, and strong model realism. Combined with clothing structure design technology, with the help of cutting knowledge, a method from 2D template to 3D dress effect display is proposed: Constructing "virtual capsids" to generate middleware, and through the correspondence between the positioning lines on the two-dimensional template and the corresponding positions of the shell, the mapping of the clothing piece to the avatar and the stitching of the virtual clothing piece are realized.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Authors' Contributions

Xiangyi Xue contributed to the writing of the manuscript and data analysis. Xiangying Xue supervised the work and designed the study. All the authors have read and agreed the final version to be published. Xiangying Xue provided great help in the revision of the final draft and agreed to be included in the author list of the article. All unanimously agreed to the above arrangement. All the authors have read and agreed the final version to be published.

Acknowledgments

This study was sponsored by Xi'an Polytechnic University.

References

[1] M. X. Li, S. Q. Yu, W. Zhang, H. Zhou, X. Xu, and T. W. Qian, "Segmentation of retinal fluid based on deep learning: application of three-dimensional fully convolutional neural networks in optical coherence tomography images," *International Journal of Ophthalmology*, vol. 12, no. 6, pp. 1012–1020, 2019.

[2] S. Pandey, N. K. Nagwani, and S. Verma, "Analysis and design of high performance deep learning algorithm: convolutional neural networks," *International Journal of Engineering Trends and Technology*, vol. 69, no. 6, pp. 216–224, 2021.

[3] Z. Luo, L. Liu, J. Yin, Y. Li, and Z. Wu, "Deep learning of graphs with ngram convolutional neural networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2125–2139, 2017.

[4] L. Huang, J. Li, H. Hao, and X. Li, "Micro-seismic event detection and location in underground mines by using convolutional neural networks (cnn) and deep learning," *Tunnelling and Underground Space Technology*, vol. 81, no. NOV, pp. 265–276, 2018.

[5] X. Ma, N. Kittikunakorn, B. Sorman et al., "Deep learning convolutional neural networks for pharmaceutical tablet defect detection," *Microscopy and Microanalysis*, vol. 26, no. S2, pp. 1606–1609, 2020.

[6] V. Puzyrev, "Deep learning electromagnetic inversion with convolutional neural networks," *Geophysical Journal International*, vol. 218, no. 2, pp. 817–832, 2019.

[7] H. Chen, Y. Song, and X. Li, "A deep learning framework for identifying children with adhd using an eeg-based brain network," *Neurocomputing*, vol. 356, no. SEP.3, pp. 83–96, 2019.

[8] A. Anton, N. F. Nissa, A. Janiati, N. Cahya, and P. Astuti, "Application of deep learning using convolutional neural network (cnn) method for women's skin classification," *Scientific Journal of Informatics*, vol. 8, no. 1, pp. 144–153, 2021.

[9] C. Shan, X. Guo, and J. Ou, "Residual learning of deep convolutional neural networks for image denoising," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 2, pp. 2809–2818, 2019.

[10] S. Albelwi and A. Mahmood, "A framework for designing the architectures of deep convolutional neural networks," *Entropy*, vol. 19, no. 6, p. 242, 2017.

[11] Q. Ni, J. Kang, M. Tang, Y. Liu, and Y. Zou, "Learning epidemic threshold in complex networks by convolutional neural network," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 11, Article ID 113106, 2019.

[12] V. Bangari, B. A. Marquez, H. Miller et al., "Digital electronics and analog photonics for convolutional neural networks (deap-cnns)," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–13, 2020.

[13] S. Luo, A. Peng, H. Zeng, X. Kang, and L. Liu, "Deep residual learning using data augmentation for median filtering forensics of digital images," *IEEE Access*, vol. 7, no. 99, p. 1, 2019.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9818912, 1 page
https://doi.org/10.1155/2023/9818912

Retraction

Retracted: Optimization and Application of Random Forest Algorithm for Applied Mathematics Specialty

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Li, "Optimization and Application of Random Forest Algorithm for Applied Mathematics Specialty," *Security and Communication Networks*, vol. 2022, Article ID 1131994, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1131994, 9 pages
https://doi.org/10.1155/2022/1131994

Research Article

Optimization and Application of Random Forest Algorithm for Applied Mathematics Specialty

Wei Li

Department of Marx School (Basic Teaching Department), Chongqing City Vocational College, Chongqing 402160, Yongchuan, China

Correspondence should be addressed to Wei Li; liwei2022215@163.com

Received 3 March 2022; Revised 19 April 2022; Accepted 27 April 2022; Published 21 May 2022

Academic Editor: Chin-Ling Chen

For unbalanced data classification, RF (Random forest) algorithm will cause problems such as poor classification performance and a large DT scale. With the advent of the era of big data, RF algorithms should have the ability to process large-scale data. Aiming at the problem that RF cannot handle unbalanced data well, this paper improves the feature selection method built in RF and proposes a new feature selection algorithm. On the basis of feature importance ranking, randomness is introduced to ensure the strength of each tree and reduce the correlation between trees. In the extended transform data set, the sensitivity of the RF model has exceeded 0.8, and that of other models has increased to about 0.65. The prediction accuracy of the centralized RF model for the company's credit rating reached 100%, while the CART model misjudged companies C6 and C7, while the Logit model misjudged companies C3, C5, and C8. Experiments prove the extrapolation of the RF model and its excellent prediction ability. In the practical application of applied mathematics specialty, the RF optimization algorithm proposed in this study can well handle continuous variables and improve the classification accuracy of RF. This paper holds that the advantages of the RF algorithm in data processing and model performance will make it more widely used in the field of enterprise credit risk evaluation.

1. Introduction

Supervised learning in machine learning algorithms is nothing more than solving classification problems and regression problems, among which there are many algorithms to solve classification problems, such as NB (Naive Bayesian), SVM (Support Vector Machine), DT (Decision Tree), and so on [1–3]. Obviously, these are all single classifiers, which are prone to overfitting problems, and there will be bottlenecks when improving their performance, so the ensemble learning algorithm came into being. Bayesian and DT are more representative of single classifier technology. These algorithms have promoted the development of classification technology to a certain extent, and all aspects of research and application have been comprehensively carried out. However, due to its own limitations, the performance improvement of a single classifier has reached an insurmountable bottleneck, so people began to put forward the idea of a multiclassifier combination [3]. Multiclassifier combination uses multiple base classifiers for classification and integrates all classification results to form a final result. RF (Random forest) is a multiclassifier combination produced under this background.

As a major direction in data mining, classification technology is a supervised machine learning method. It trains the training set to get the learner model and then tests the test set with this model to get the classification result. Rather et al. proposed a method to predict the activity of cannabinoid receptor agonists using RF technology [4]. BZBA and others introduced the Bagging method and systematically expounded the RF algorithm, and the RF algorithm officially became an important part of the data mining classification algorithm [5]. Vassallo et al. used the RF algorithm to study the land coverage area and found that the RF algorithm can train faster than other combination algorithms [6]. Cao et al. also applied RF to time series to detect the change points of time series [7]. Although the performance of the RF algorithm has been gradually improved, and the scenes used are more and more extensive, there are still some defects in some aspects, such as feature selection and processing of unbalanced data sets. At home, researchers focus on the application of the RF algorithm but there is not much research on the optimization and improvement of the RF algorithm.

Because the RF algorithm has the problem of a low prediction rate of subcategories when dealing with unbalanced data sets, researchers optimized the data preprocessing process of the RF algorithm. They calculate the attribute weights, calculate the correlation between attributes according to the Chi-square test, and calculate by analyzing the correlation between each attribute and the target attribute. Finally, they sort the weights of each attribute, and in the process of feature selection, they tend to the attribute ranked first, which enhances the DT intensity of each tree and reduces the correlation coefficient between trees, thus improving the classification accuracy of the RF algorithm [8]. In this paper, on the basis of full access to relevant information at home and abroad, aiming at the problems of RF in theory and application, the optimization of RF algorithm and application for applied mathematics major is mainly carried out.

The main innovations of this paper are as follows:

(1) In the aspect of RF self-optimization, the influencing factors of classification performance of RF algorithm are analyzed in detail. Aiming at the phenomenon of different RF performances caused by different node splitting algorithms during RF generation, an RF optimization algorithm based on linear transformation is proposed.

(2) After the above optimization, this paper actively explores the application of the optimized RF algorithm in enterprise credit risk evaluation and constructs a six-level risk evaluation model based on RF. By comparing the models, it proves that RF has excellent stability, extrapolation, and predictive ability.

The paper is divided into five chapters.

The first chapter introduces the research background and outlines the main tasks of this paper. The second chapter introduces the research status of the RF algorithm. The third chapter introduces the optimization and application of the RF algorithm. The fourth chapter compares the performance of this model through experiments. The fifth chapter is the full-text summary.

2. Related Work

2.1. Optimization Method of RF Algorithm. Xu et al. proposed the random survival forest algorithm and introduced the concept of survival trees in RF [9]. In the process of the bootstrap resampling method, the algorithm has to generate a corresponding analysis tree for each training subset formed by sampling. Hao et al. put forward the quasiadaptive classification RF algorithm. First, it was found that the Adaboost algorithm has great advantages in adaptive self-help sampling weight and adaptive voting weight setting [10]. Fornaser et al. mixed the C4.5DT algorithm, and Fornaser et al. mixed the C4.5DTalgorithm and CART (Classification and Regression Tree) algorithm into one algorithm and used the mixed algorithm to generate the RF algorithm, which improved the accuracy of RF [11]. Paul et al. proposed an improved RF algorithm based on DT clustering to extract DTs with low classification accuracy and high similarity. Experiments show that this algorithm is higher than the traditional RF algorithm in integration accuracy and classification efficiency [12]. Li introduced the theory of quantile regression, applied quantile regression to the process of DT generation and decision-making, and proposed the quantile regression forest algorithm. He mathematically proved the consistency of quantile regression to the forest [13]. Chen et al. developed a new cost-sensitive RF algorithm from the perspective of applying a cost-sensitive learning algorithm to solve the classification problem of unbalanced data sets [14]. Dou et al. proposed a new RF feature selection algorithm by analyzing the relationship between the intensity and correlation coefficient of each tree in RF [15]. The main idea of this algorithm is that by analyzing the upper bound of RF generalization error, it is found that increasing the intensity of DT in the forest can reduce the generalization error of RF. Santra et al. proposed an improved RF classifier, which is classified by the minimum number of trees and limited the number of DTs in RF according to the importance of features. Experiments on different data sets show that the classification error is significantly reduced [16].

2.2. Application of RF Algorithm. RF algorithm is widely used in many fields because of its good comprehensive performance. Asadi et al. applied the RF algorithm to environmental protection, used it to predict urban smog, and finally analyzed and expounded the control measures of smog [17]. Prasad et al. established the fund rating model by using RF algorithm and thought that the information ratio was the most important index of fund evaluation, followed by a determinable coefficient. The research proved that the stability and accuracy of the model reached an excellent level [18]. Kwan et al. introduced the nonparametric RF method into the field of fund excess return direction prediction in China, which proved that the RF method was superior to random walk and support vector machine algorithms, and also proved the predictability of the domestic financial market to a certain extent [19]. Wu et al. used the RF algorithm to conduct on-site monitoring and penetration prediction of plasma arc welding [20]. Mei et al. studied the identification of commuters based on RF of smart card data and compared it with the discriminant analysis method [21].

3. Methodology

3.1. Overview of RF Algorithm. On the basis of constructing bagging integration with DT-based learners in RF, the selection of random attributes is further introduced into the training process of DT. RF algorithm is simple, easy to implement, low in computational cost, and shows strong performance in many practical tasks. Therefore, the RF method is an extension of the traditional DT method, which combines multiple DTs to improve prediction accuracy.

DT is a typical single classifier. To use it for classification, we need to build a DT model based on training data and then use this model to classify unknown sample data. The pruning process is to prune some subtrees or leaf nodes in the DT model, and the main purpose is to avoid overfitting by simplifying the DT model. Firstly, according to the selected feature evaluation criteria, child nodes are recursively generated from the root node from top to bottom until the leaf node is reached.

In the process of DT node splitting, ID3 algorithm takes the information gain of features as the feature evaluation standard, the feature with the largest information gain as the test attribute, and the calculation of information gain is based on information entropy. Let X be a discrete random variable with finite values, and its probability distribution is as follows:

$$P(X = x_i) = p_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

Then the entropy of the random variable X is defined as follows:

$$H(X) = -\sum_{i=1}^{n} p_i \log\, p_i. \tag{2}$$

The generalization error of DT in all forests converges to the following expression:

$$\underset{n\longrightarrow\infty}{lin}\, PE^* = P_{xy}\left(P_\Theta (k(X,\Theta) = Y) - \max_{j \neq Y} P_\Theta (k(X,\Theta) = J) < 0 \right), \tag{3}$$

where n is the number of trees in the forest.

The basic idea of RF classification: Firstly, k samples are extracted from the original training set by bootstrap sampling, and the sample capacity of each sample is the same as that of the original training set; Then, k DT models are established for k samples, and k classification results are obtained. Finally, according to the k classification results, vote on each record to determine its final classification. The schematic diagram is shown in Figure 1.

In the process of generating the RF algorithm, bagging sampling technology is mainly used to generate training subsets from the original training set. The size of each training subset is about two-thirds of that of the original training set, and each sampling is random and put back to sampling, which makes the samples in the training subset have certain duplication, and the purpose of this is to prevent DT in the forest from generating local optimal solutions.

Using Bagging to build RF has two meanings. On the one hand, it can improve the classification accuracy of the RF algorithm. Because the samples are put back, almost 37% of the samples are not in the training subset, which can prevent abnormal data and noise data from appearing in the training subset to a certain extent, and can get higher performance

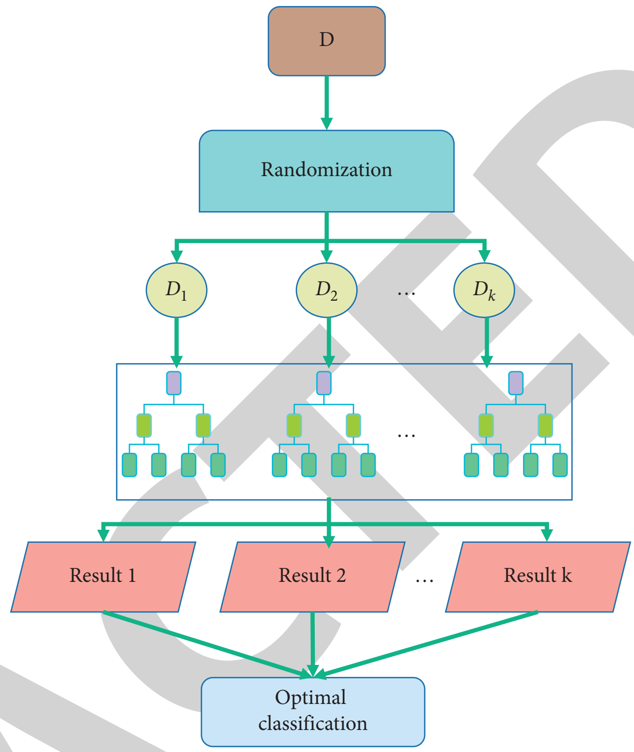


Figure 1: Schematic diagram of RF classification.

DT compared with the original data set. A random feature vector can reduce the correlation between trees in forest, thus reducing overfitting and improving the classification accuracy of forest, which introduces another random factor for RF.

In RF, if there are continuous variables, it is common practice to divide the values of these continuous variables into different intervals, that is, "discretization." However, due to the great relationship between the algorithm complexity after discretization and the reduction rate of the data set, it takes a lot of time to analyze and calculate the node splitting standard, which greatly affects the execution speed of the algorithm. Therefore, the discretization of continuous variables is content that needs to be optimized in the RF algorithm.

3.2. Optimization Method of Algorithm Based on. Generally speaking, high-dimensional data is relative to traditional low-dimensional data, and the number of attributes can often reach hundreds of thousands or even higher. Mining algorithm for high-dimensional data has always been a hot research topic, and the classification of high-dimensional data is a difficult pattern recognition problem. Traditional classification algorithms have problems such as low classification accuracy, easy overfitting, and long-running time when processing high-dimensional data. In this chapter, aiming at the problems of low classification accuracy and large generalization error of RF algorithm in high-dimensional data classification, an intelligent algorithm-based RF feature selection and parameter optimization for high-dimensional data is proposed.

Feature selection can be described as a process in which an optimal feature subset $T' = \{t_{1'}, t_{2'}, \ldots t_{s'}\}$ is selected from the feature set $T = \{t_1, t_2, \ldots t_s\}$ and T' can contain most of the information of the original sample, among which $s' < s$, the purpose of feature selection is to make the classification or regression model constructed by feature subset T' achieve similar or even better prediction accuracy than before feature selection.

Therefore, according to the above two methods, we can study how to improve the treatment method of the class imbalance problem. The first scheme is to study the data distribution using the balanced RF method, which can be easily integrated into RF. Another approach is to apply the generation-sensitive algorithm to RF. This method can be accomplished using a weighted RF algorithm.

Smote (Synthetic Minority Over-sampling Technique) algorithm improves the random upsampling method. Because random upsampling is a random replication of negative samples, and many of the new data sets generated are duplicated, it is difficult to effectively solve the data imbalance problem.

SMOTE algorithm first looks for k nearest negative samples around each negative sample and then constructs a new negative sample between this sample and k adjacent samples. The process of interpolation synthesis is shown in the following formula:

$$P_{ij} = x_i + \mathrm{rand}(0,1) \times \left(y_{ij} - x_i\right), \tag{4}$$

$x_i\,(i = 1, 2, \ldots, n)$ is negative samples, and n represents the number of negative samples; $y_{ij}\,(j = 1, 2, \ldots, m)$ is the m nearest neighbor samples adjacent to x_i; P_{ij} represents a new sample synthesized by sample x_i; $\mathrm{rand}(0,1)$ represents any random number between (0, 1).

However, SMOTE algorithm has two problems: First, when choosing the nearest neighbor, there is certain blindness about how much k value to take. How many nearest neighbor samples to take need to be solved by users themselves? Sometimes, we have to repeatedly test according to specific data sets, and we have to explore what kind of k value makes the algorithm optimal.

In this paper, a new discretization algorithm of continuous variables based on x^2 correction is designed, and its calculation process is as follows:

Calculate the number k of decision attributes in two adjacent intervals of a certain attribute value, and calculate the theoretical times E_{ij}, and the formula is as follows:

$$E_{ij} = R_i \times \frac{C_j}{N}. \tag{5}$$

In which: $R_i = \sum_{j=1}^{k} A_{ij}$ is the number of samples in the i interval; $C_j = \sum_{i=1}^{k} A_{ij}$ is the number of class j samples; N is the total number of samples in two adjacent intervals.

If the theoretical degree E_{ij} of a certain group is less than 5, it should be combined with its adjacent group or groups until the theoretical degree E is greater than 5 or there is only one set of data in an interval, and then the k value should be recalculated.

The value of the x^2 statistic is calculated in two cases: When $k < 2$, the formula of x^2 statistics is as follows:

$$x^2 = \sum_{i=1}^{2}\sum_{j=1}^{k} \frac{\left(\left|A_{ij} - E_{ij}\right| - 0.5\right)^2}{E_{ij}}, \quad j = 1; i = 1, 2. \tag{6}$$

When $k \geq 2$, the formula of x^2 statistics is as follows:

$$x^2 = \sum_{i=1}^{2}\sum_{j=1}^{k} \frac{\left(A_{ij} - E_{ij}\right)^2}{E_{ij}} \quad j = 1, 2, \ldots, k; i = 1, 2, \tag{7}$$

where k is the number of target variable categories; i number two adjacent intervals; A_{ij} is the number of class j samples in the i th interval in two adjacent intervals;

The order of merging intervals is determined by D value, and its formula is as follows:

$$D = \frac{x_a^2 - x^2}{\sqrt{2v}}. \tag{8}$$

Select the interval with the smallest D value to merge. After all the continuous attribute variables in the data set are subjected to the above steps, the reduction process of the data set is completed. That is, the discretization of the continuous variables is realized. The program flow of the continuous variable discretization algorithm based on x^2 correction is shown in Figure 2 below.

3.3. Application of Optimized RF Algorithm.

3.3. Application of Optimized RF Algorithm. Credit means that one party obtains something from the other party and promises to repay it in the future. However, in actual transactions, it is often impossible to ensure whether the debtor has sufficient repayment ability and willingness. Therefore, creditors can only evaluate the debtor's repayment ability and willingness in probability, and the uncertainty in this decision-making process will lead to credit risk.

The characteristics of credit risk mainly reflect four aspects:

(1) Credit risk is systematic. Systemic risk is the inherent risk of the financial market, which is closely related to macroeconomic changes.

(2) The formation of credit risk is related to people's subjective will. In credit transaction activities, when one party has obvious floating losses, in order to ensure personal economic interests, strategic default is often chosen.

(3) The profit and loss caused by credit risk is usually asymmetric. In the stock market, the rise and fall of the stock price can be considered symmetrical, so the profits and losses brought by it are also symmetrical in theory.

(4) The formation of credit risk is a cumulative process. Looking at the origin of credit risk events, most of them are related to the deteriorating historical business conditions. Local credit risk is transmitted to the whole financial market, causing a chain reaction and eventually even causing the disorder in the whole financial system.

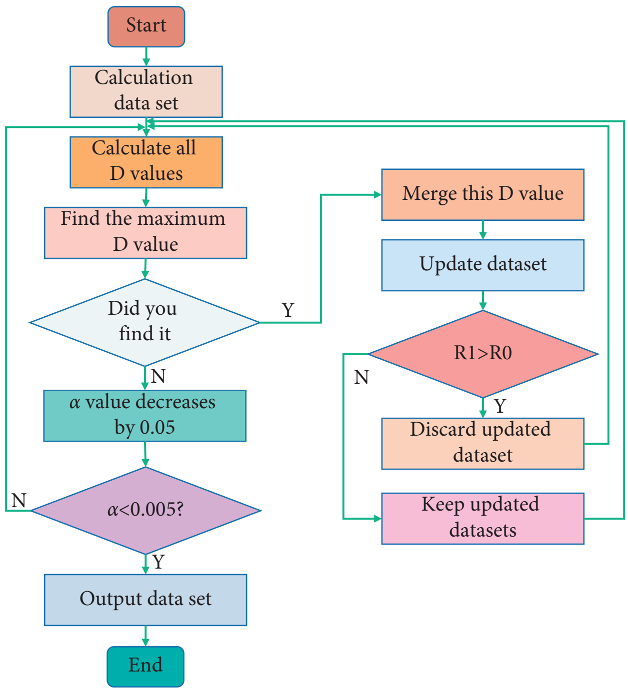


FIGURE 2: Algorithm program flow chart.

RF algorithm is an integrated algorithm. CART and Bagging methods are combined in the classifier, which is more adaptive than other data mining algorithms. As a nonlinear modeling tool, RF can well deal with classification problems with few rules constraints and missing data, and it is also effectively used in the credit approval of commercial banks. From the past research, the accuracy of RF algorithm is better than other algorithms such as neural network algorithm, which is also the reason for further research in this paper.

This chapter will conduct an empirical study with the manufacturing industry as the research background and corporate credit risk as to the research object. Choosing and constructing an evaluation index system is an important step of risk evaluation research. Choosing indicators to build a credit evaluation index system must be based on the purpose of evaluation, carefully analyze the things to be inspected, find out the factors that affect the evaluation objects, select some main factors from them, and build a credit evaluation index system. In this paper, 26 quantitative indicators including profitability, cash flow capacity, operational capacity, development capacity, short-term solvency, and long-term solvency and a qualitative indicator of management quality are selected.

This paper assumes that the financial statements published online by the eight companies involved in the research are consistent with those provided to credit rating companies and that the data are true and reliable. This paper introduces pseudo data. The pseudo data satisfies the following inequality conditions:

$$x(1-w) \leq x' \leq x(1+w). \tag{9}$$

In which: x is the original real data; w is the width of the random number; x' is false data.

Let DT have M leaf nodes, $R_m\,(m = 1, 2, \ldots, M)$ is the decision area under the m-th leaf node, and C_m (constant) is the decision value, which indicates the proportion of target classification under the decision area.

According to the process of DT discrimination, the new sample will eventually fall into one of the decision areas after top-down judgment, and the probability that the sample belongs to the target class is expressed by C_m. Thereby having the following leaf node discrimination function:

$$fc(X) = \sum_{m=1}^{M} C_m I(X, R_m). \tag{10}$$

In which, X is the input vector and $I(\cdot)$ is an illustrative function, which means that when the discrimination of X falls into the region R_m, the value is 1; otherwise, it is 0.

Choosing the feature-based credit risk evaluation model means choosing the best subset of indicators from the candidate evaluation index system as the evaluation index system of the model. In the traditional sense, feature selection refers to the self-correlation analysis between features to get the evaluation index system by removing the features with high linear correlation. Because the Wrapper algorithm has advantages over specific algorithms, this paper uses the Wrapper method for feature selection, and the specific process is shown in Figure 3:

The specific steps of selecting the optimal feature subset by RF are as follows:

(1) Randomly select a certain percentage of data from the sample data as training data.

(2) Use the package VarSelRF in R language to calculate the importance of each evaluation index in training data and sort it.

(3) Carry out many experiments, compare the OOB error rate of each experiment, evaluate the influence of the number of indexes on the model performance, and take the index set with the lowest OOB error rate as the index set for feature selection.

When inputting model data, it is usually necessary to normalize the data. The advantage of normalization is that it can eliminate the dimensional influence of different data. When the data is not normalized, the larger evaluation index may weaken the influence of the smaller evaluation index on the model. The most commonly used normalization method is the maximum-minimum method. Generally, we classify the index as between [−1, 1] and [0, 1].

Because some issuers in private placement bonds did not disclose financial information, the corresponding indicators were missing, so these samples were excluded, and the sample distribution is shown in Table 1.

After SMOTE expansion of the data set, consider grouping the continuous variables to reduce the dimension. Among the 28 variables selected, there are only two values of x^2 (enterprise nature), so there is no need to group them again. For the other 27 continuous variables, most of the values range from positive infinity to negative infinity.

4. Experiment and Results

In order to study the parallel RF based on the MapReduce model, this experiment selects Ionosphere, Crowdsourced

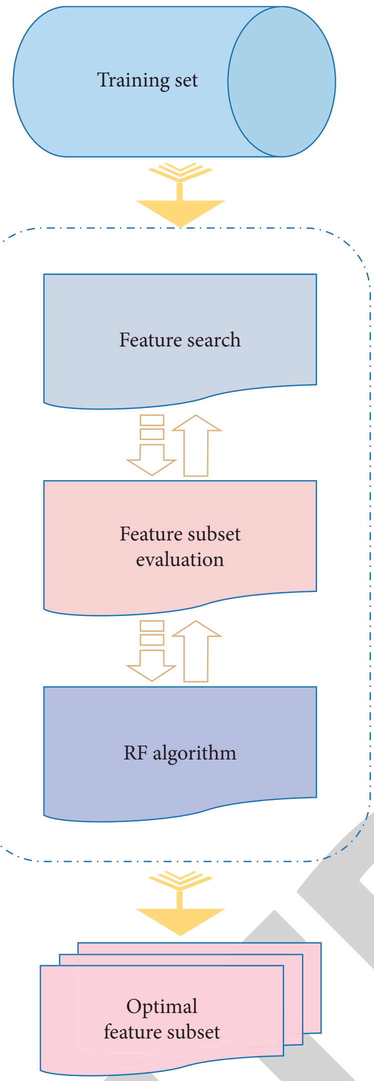


Figure 3: Feature selection process.

Table 1: Original sample distribution.

Credit event	Number of positive samples	Negative sample number	Total
Guarantor compensation	0	2	2
Guarantor's rating upgrade	147	0	147
Main body rating upgrade	10	0	10
Break a contract	0	22	22
Subject rating maintenance	206	0	206
Downgrading of subject	0	77	77

Mapping, KDDTest, and Covertype datasets in the UCI machine learning database, and the number and size of the datasets increase in turn. The data set description is shown in Table 2.

In the RF model, the number of DTs is uniformly set to 100, and the running time and acceleration of stand-alone RF and parallelized RF on Ionosphere, Crowdsourced Mapping, KDDTest, and Covertype data sets are as shown in Figures 4 and 5. "node" represents the number of data nodes.

Table 2: Data set description.

Serial number	Dataset name	Number of samples	Number of attributes	Data set size
1	Ionosphere	350	33	0.0771
2	Crowdsourced mapping	10553	29	1.12 M
3	KDDTest	125677	40	3.36 M
4	Covertype	590161	55	70.82 M

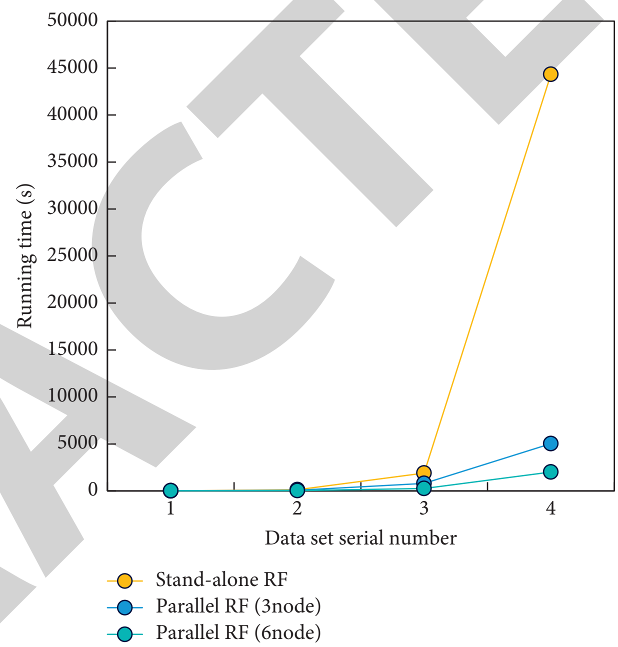


Figure 4: Run time comparison.

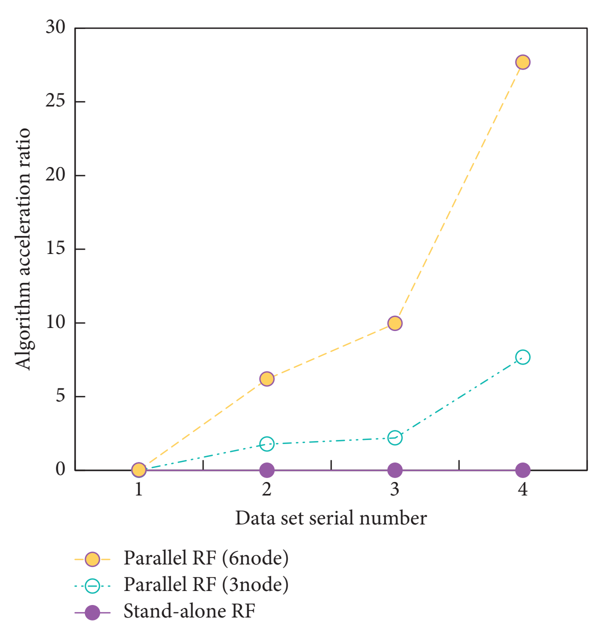


Figure 5: Comparison of algorithm speedup ratio.

It can be seen that the speedup ratio of the parallel RF algorithm based on the MapReduce model in a large-scale data environment has increased on each data set, and the speedup ratio is larger with the larger data set, and the speedup ratio is larger with more nodes.

Comparing the running time of parallel RF with that of stand-alone RF, we can see that with the increase of data set size, the running time of stand-alone RF will increase rapidly, while that of parallel RF will increase very slowly, and with the increase of node number, the running time of RF algorithm will drop more significantly. This shows that the parallel RF algorithm can better adapt to large-scale and massive data, and the efficiency of the algorithm will be higher.

When RF algorithm selects features, the Gini index is used as the measurement standard, and the random forest contains many DTs. In the process of establishing DTs, it is usually necessary to calculate the corresponding information bribe of each feature, that is, the reduction of impurity. The information gain is to calculate the information bribe of the class label on the whole data set and then calculate the difference according to the information bribe of each feature.

As the feature is selected, the impurity will be greatly reduced, so the importance of other associated features will be greatly reduced, and it is difficult for other associated features to be selected again. Thus, the initially selected features are very important, while the other associated features are of low importance. This phenomenon reduces the importance of other features, but in fact, the importance of these features is similar.

When selecting features, on the one hand, select important features according to the Gini index; on the other hand, select features in a certain proportion in two-interval features, respectively, and do partial random, so as to balance the strength and correlation of features.

Here, we use the ratio of 3 : 7 to segment the data, randomly extract 90% of the features in the top 30% of importance, and randomly extract 60% of the features in the bottom 70% of importance so as to ensure that the extracted features account for about 70% of the total features and build a DT, as shown in Figure 6.

The time for RF modeling and prediction is relatively fast, and the time cost for evaluating features by chi-square test is also very small, so this method can ensure the running efficiency of the model.

The transformation of data sets has little influence on the importance of variables. Next, we will further explore the change in model performance before and after the expansion transformation. Set the original data to 7 : 3 according to the distribution ratio of the training set and test set and establish three models of RF, DT, and logit, respectively. After 100 simulations, the comparison results of the average accuracy of each model are shown in Table 3.

From the analysis in Table 3, it can be seen that the extended transformation of the data set is beneficial in improving the accuracy of the model. The reason is that the extended transformation of data balances the structure of sample data and makes the performance of the classifier more stable. At the same time, continuous variables are discretized to avoid the influence of extreme values to a certain extent, which makes the model more adaptable. Therefore, the extended transformation optimization measures for data sets improve the overall accuracy of each model.

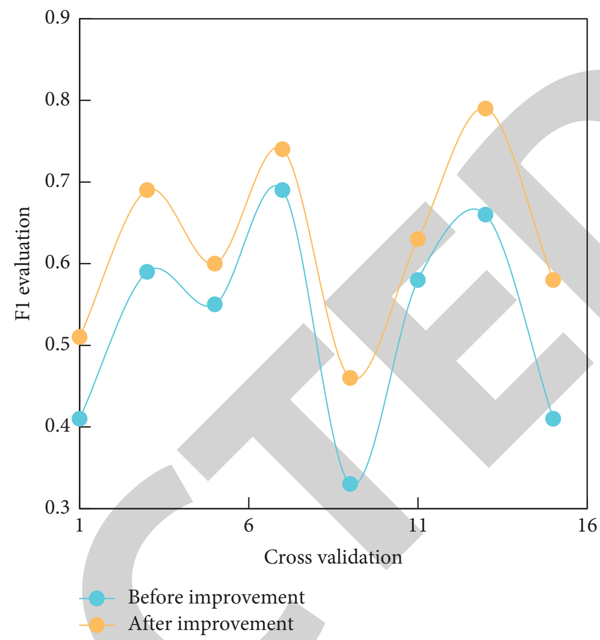


Figure 6: Comparison of results before and after feature improvement.

ROC curve takes FPR (false positive rate, the classifier mistakenly judges the actual normal as the number of default accounts for the actual total normal of the sample) as the abscissa axis and TPR (sensitivity, the classifier predicts the default accounts for the total number of actual defaults, that is, the coverage rate of default discrimination in the sample) as the ordinate axis. The ROC curve of each model is shown in Figure 7.

In the extended transformed data set, the sensitivity of the RF model has exceeded 0.8, and the sensitivity of other models has increased to about 0.65. This shows that under a certain risk tolerance, the accuracy of the model has also been greatly improved, which is particularly important in determining the appropriate risk preference.

In order to confirm the robustness and extrapolation of the RF model more reliably, this paper makes several experiments on the forecast set (the forecast set refers to the eight companies selected for forecasting in this paper) by using the RF model, CART model, and Logit regression model respectively, and the experimental results are shown in Figure 8:

From Figure 8, it can be seen that the prediction accuracy of the RF model for the company's credit rating in the prediction set reached 100%, while the CART model misjudged company C6 and C7, while the Logit model misjudged companies C3, C5, and C8. The experiment further proved the extrapolation of the RF model and the excellent prediction ability of the RF model.

Table 3: Comparison of average accuracy of models under different data sets (%).

Contrast model	Training set			Test set		
	Positive sample accuracy	Negative class sample accuracy	Overall accuracy	Positive sample accuracy	Negative class sample accuracy	Overall accuracy
RF	95.36	90.01	93.36	91.96	86.54	88.17
CART	90.06	89.77	89.06	88.91	80.24	81.14
Logit	80.13	80.09	77.86	77.96	75.23	79.64
Optimize RF	97.63	95.52	96.17	92.33	91.69	94.98

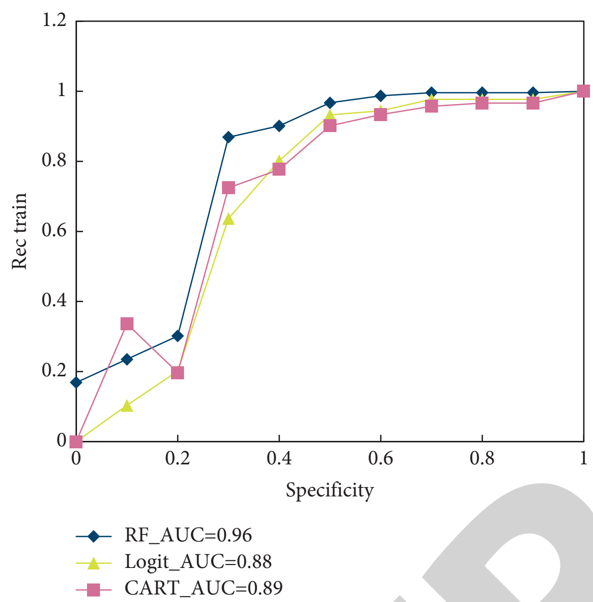


Figure 7: ROC curve based on extended transformation data set.

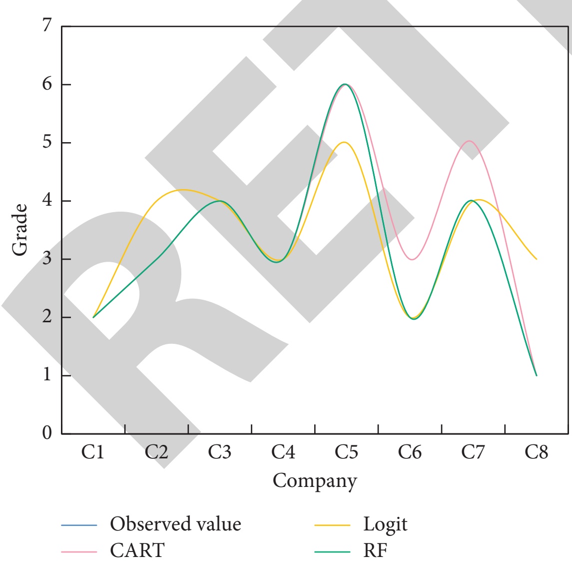


Figure 8: Comparative analysis of prediction ability of models.

5. Conclusions

RF algorithm is an algorithm with high classification accuracy and high efficiency, and its theory and method research have been mature, and it has been applied in many fields with good results. In the aspect of the optimization of unbalanced data sets, this paper proposes a new algorithm to solve the unbalanced problem, which better solves the unbalanced problem of data sets and significantly improves the classification performance of the RF algorithm on unbalanced data sets. The experimental results show that the parallel RF algorithm has greatly improved the processing ability of large-scale data. CART and Logit are used as experimental reference models, which prove that the stability, extrapolation, and prediction ability of the RF model are far superior to the experimental reference model, and there is no overfitting phenomenon in the experimental process. It is proved that RF has better performance when it is used to build the credit risk evaluation model of listed companies in the production industry.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

References

[1] M. Kang, S. K. Gonugondla, S. Lim, and N. R. Shanbhag, "A 19.4-nJ/Decision, 364-K decisions/s, in-memory random forest multi-class inference accelerator," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 7, pp. 2126–2135, 2018.

[2] S. Ma, M. Chen, J. Wu, Y. Wang, B. Jia, and Y. Jiang, "Intelligent fault diagnosis of HVCB with feature space optimization-based random forest," *Sensors*, vol. 18, no. 4, p. 1221, 2018.

[3] Z. Tang, Z. Mei, W. Liu, and Y. Xia, "Identification of the key factors affecting Chinese carbon intensity and their historical trends using random forest algorithm," *Journal of Geographical Sciences*, vol. 30, no. 5, pp. 743–756, 2020.

[4] T. A. Rather, S. Kumar, and J. A. Khan, "Multi-scale habitat modelling and predicting change in the distribution of tiger and leopard using random forest algorithm," *Scientific Reports*, vol. 10, no. 1, Article ID 11473, 2020.

[5] Z. Bei, Z. Yu, N. Luo, C. Jiang, C. Xu, and S. Feng, "Configuring in-memory cluster computing using random forest," *Future Generation Computer Systems*, vol. 79, pp. 1–15, 2018.

[6] D. Vassallo, R. Krishnamurthy, T. Sherman, and H. J. S. Fernando, "Analysis of random forest modeling strategies for multi-step wind speed forecasting," *Energies*, vol. 13, no. 20, p. 5488, 2020.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9825634, 1 page
https://doi.org/10.1155/2023/9825634

Retraction

Retracted: Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology

Security and Communication Networks

Received 8 December 2022; Accepted 8 December 2022; Published 22 January 2023

Security and Communication Networks has retracted the article titled "Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] Z. Wang and G. Wang, "Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology," *Security and Communication Networks*, vol. 2022, Article ID 4912841, 11 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4912841, 11 pages
https://doi.org/10.1155/2022/4912841

WILEY | Hindawi

Research Article

Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology

Zhangfu Wang[1] and Guohu Wang[2]

[1]*Quzhou College of Technology, Quzhou, Zhejiang 324000, China*
[2]*Jing Hengyi School of Education, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China*

Correspondence should be addressed to Guohu Wang; wgh@hznu.edu.cn

Received 11 March 2022; Revised 29 March 2022; Accepted 13 April 2022; Published 21 May 2022

Academic Editor: Chin-Ling Chen

In order to improve the teaching effect of intelligent education, this paper combines machine learning and fuzzy algorithm to improve the application effect of educational information technology and the use of information platform. Moreover, this paper collects the user's behavior information in the process of using the platform as the user's interest description and selects personalized upgrade resources for the user. In addition, this paper introduces the similarity as a personalized recommendation evaluation standard and calculates the average similarity between the recommended related resources and the user's interest. Finally, this paper constructs an educational information system based on improved machine learning and fuzzy algorithms. The experimental research results show that the intelligent system proposed in this paper can play an important role in educational information processing and teaching quality improvement.

1. Introduction

The application of educational information technology refers to the educational and teaching activities that use educational information technology under the guidance of educational science and information science theory and in accordance with certain educational goals, educational principles, and educational plans. Through the analysis of the basic concepts of educational information technology and its system structure, we can find that the application of electronic education or modern science and technology in education and teaching that we have been engaged in for a long time is the application of educational information technology.

There is another reason for the concept of educational information technology, that is, educational informatization is the process and result of the popularization and application of educational information technology. This is just as networking is the promotion and application of network technology, and multimedia is the promotion and application of multimedia technology [1]. The "five modernizations" that represent the basic characteristics of educational informatization, namely, multimedia display of educational information, digitization of educational information processing, CD-based educational information storage, networked educational information transmission, and intelligent educational information management are the processes and results of the promotion and application of various educational information technologies. From the perspective of informatization, it can be said that educational information technology is a general term for various technologies used to realize educational informatization. The purpose of the national informatization construction spending huge sums of money to carry out some informatization projects is to promote the application of information technology, make it a multiplier of productivity, and produce advanced and efficient social functions [2].

In the era of "Internet + education", the integration of information technology and classroom teaching is an inevitable trend of education reform. At present, traditional information technology cannot cope with the challenges brought by the explosion of information in the digital environment to classroom teaching, and educational information technology is gradually playing a role. The actual process of educational information technology facing the teaching of various disciplines focuses on the development and application of information technology suitable for teaching [3].

This paper combines improved machine learning and fuzzy algorithms to explore the application of educational information technology, improve the application effect of educational information technology, and provide a reference for the further development of subsequent educational information technology.

2. Related Work

Traditional education has always been faced with practical problems such as lack of learning resources, single learning method, and single teaching mode. Of course, this is not unrelated to the ideal education that people have been looking forward to, that is, learning can be more convenient, relaxed, and free. The invention of technology contains people's purposes and values. The application of technology in the field of education is expected to produce beneficial value and promote practical teaching. In this way, it is logical for some scholars to study the educational application of technology [4]. In recent years, with the gradual development of educational technology, the exposure of various problems in educational technology practice, and the frequent occurrence of alienation, some researchers have begun to consciously think about the nature of educational technology and human development, philosophical methodology has gradually introduced into the study of educational technology, and Marx's theory of alienation and his thoughts on the all-round development of human beings are gradually cited by people and they try to find solutions to problems from Marx's theoretical nutrition [5]. This undoubtedly promotes the deepening of the research on the basic theory of educational technology. The opportunities and challenges brought by modern information technology to education are unprecedented. It is an inevitable choice to reshape the form of education under the background of information technology by extensively absorbing the research results of the philosophy of technology and re-examining the problem of alienation in the field of education [6]. The problem of technology alienation is a very important issue in the philosophy of technology. The problem of modern educational technology alienation discussed in the literature [7] undoubtedly draws theoretical nourishment from the philosophy of technology and regards the development of technology, education, and human beings as the starting point and destination of this paper. In the following, the author will make a general summary of the research on this issue at home and abroad.

Literature [8] raised the duality of computer and online education and believed that the purpose of developing technology in education should be aimed at people and should be filled with care for people's development. However, one-sided emphasis on the practical value of technology in educational technology will, to some extent, lead to indifference to cultural and spiritual life other than technology, ignore the value of life, and make education lose its responsibility of "making people into people." Value pursuit: literature [9] pointed out that technology itself does not help students to learn actively, the difference lies in how technology is applied to the learning process, and the key lies in how technology is integrated to provide an effective learning experience. Literature [10] believes that when technology is applied to teaching and promotes learning, it also produces a series of new problems and points out that, among many new problems, it is particularly noteworthy that, with the continuous strengthening of technological power, technology itself has from an extended tool of human power to an ideology that controls people and educational technology has evolved into a hegemonic power in the education and teaching system, which seriously threatens the subject and object of education—teacher and student. The dominant position of human beings faces the danger of losing their free choice.

Literature [11] defines information technology as: science applied in information processing and processing, training methods, and management skills of technology and engineering; application of skills and methods; computer and its interaction with humans and machines; and corresponding social, economic, and cultural storage things. Literature [12] divides the concept of information technology into two types: chivalrous and broad. In a broad sense, information technology is a general term for technologies that can expand and extend human information organs. The basic content of information technology is composed of sensing technology, communication technology, computer technology, and control technology. The chivalrous information technology refers to the human experience of collecting, transmitting, processing, storing, publishing, and retrieving various information such as data, language, text, sound, pictures, and images which is only the sum of its means and tools. These technologies enhance the human ability to process information. The understanding of information technology from different angles is also different. In this article, it is more appropriate to use chivalrous information technology, which can promote the modernization of educational ideas, methods, concepts, and means.

Literature [13] believes that the alienation phenomenon existing in the application of modern educational technology in modern teaching is embodied in three categories: purpose without means, means inconsistent with the purpose, and means becoming the purpose. It is an effective way to eliminate alienation by improving teachers' awareness of education and enhancing their practical ability to use educational technology. Literature [14] summarizes the connotation and performance of educational technology alienation and points out the inevitability and controllability of educational technology alienation and then discusses the elimination countermeasures of educational technology alienation from the level of educators. Literature [15] has a good understanding of the modernization of education in the field of education. The source of modern information technology alienation has been deeply analyzed, and it is pointed out that the competition between technology and people is the direct source of alienation. Literature [16] discussed the problem of technological alienation from the specificity of educational practice.

3. Educational User and Resource Description Model and Modeling Method

The basic information of users is mainly described in terms of school name, school type, characteristic theme, natural situation, and informatization level. Among them, the natural situation includes the level of economic development where the school is located, the area of the school, the number of classes, the number of students, the number of teachers, and the composition of teacher education. The level of informatization includes teachers' information literacy level, students' information literacy level, campus network construction, computer room configuration, and teacher computer configuration.

In order to reflect user interest more realistically, the user interest model is represented by background and tense vector space model (BTVSM). That is, subject and school stage are introduced into the model as background constraints, and the interest weight function $\omega_n(\overline{T}_n)$ based on temporal changes is introduced into the vector space, so as to calculate the attenuation and update of the user interest weight.

At certain time t, the user interest model is expressed as

$$\begin{aligned} &UI = \{s, g, K\},\\ &\quad s \in S, \quad g \in G,\\ &K = \{(k_1, \omega_1(\overline{T}_1)), (k_2, \omega_2(\overline{T}_2)), \ldots, (k_n, \omega_n(\overline{T}_n)\},\\ &\overline{T}_n = \{t_{n1}, t_{n2}, \ldots, t_{nm}\}. \end{aligned} \tag{1}$$

Among them, S represents the subject set; G represents the school section set; K represents the user interest keyword vector space; k_n is the nth keyword describing interest; $\overline{T}_n$ represents the time set of each submission of the keyword k_n; t_{nm} is the time of the mth submission of the keyword k_n; and $\omega_n(\overline{T}_n)$ is the weight function of the keyword k_n with respect to time.

For the attenuation and update of user interest, the calculation based on the time window mechanism is adopted here, that is, within a certain time window Δt, if the keyword is submitted, the weight is increased; otherwise, the weight is attenuated.

Assumptions are as follows:

(1) If each time the keyword k_n is submitted in each time window Δt, the interest weight increases by unit a

(2) If the keyword k_n is not submitted within each time window Δt, the interest weight decays by unit b;

Then, at certain time t, the interest weight function of the keyword k_n is expressed as [17]

$$\begin{aligned} &\omega_n(\overline{T}_n) = \begin{cases} \sum_{i=1}^{f(t_{n,t})} \left(f\left(t_{nl} + (i-I)\cdot\Delta t\right)\cdot a - c\cdot b\right), & \omega_n(\overline{T}_n) > 0,\\ 0, & \text{otherwise}, \end{cases}\\ &c = \begin{cases} 0, & f\left(t_{nl} + (i-1)\cdot\Delta t\right) > 0,\\ 1, & \text{otherwise}. \end{cases} \end{aligned} \tag{2}$$

In the formula, $f(t_{n1}, t)$ represents the number of time windows Δt in the time period $[t_{n1}, t]$ and $f(t_{nl} + (i-I)\cdot\Delta t)$ represents the number of times the keywords were submitted in the time window $[t_{nl} + (i-l)\cdot\Delta t, \quad t_{nl} + i\cdot\Delta t]$.

The user interest model is described in a tree structure as shown in Figure 1.

The resource description file in the system adopts the representation method similar to the user interest model, namely background and cognitive level vector space model (BCLVSM) representation. Each resource R is represented by the vector space of the subject and academic stage as the background combined with the cognitive level of the required user [18]:

$$\begin{aligned} &R = \{s, g, cl, K\},\\ &\quad s \in S, \quad g \in G, \quad cl \in \mathrm{CL},\\ &K = \{(k_1, \omega_1), (k_2, \omega_2), \ldots, (k_n, \omega_n)\}. \end{aligned} \tag{3}$$

Among them, S represents the subject set; G represents the school section set; CL represents the cognitive level of the required user; K represents the keyword vector space of the resource description file; k_n is the keyword of the nth resource description; ω_n is the weight of the keyword k_n in the resource description; and $\sum_{j=1}^{n}\omega_j = 1$ [19].

The cognitive level CL of the users required for the resource is expressed by fuzzy language, and the domain of discourse is CL = $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$:

Fuzzy set $L = 1/1 + 0.8/2 + 0.6/3 + 0.4/2 + 0.2/5 + 0.1/6$

Fuzzy set $M = 0.7/5 + 0.8/6 + 1/7 + 0.8/8$

Fuzzy set $H = 0.7/7 + 0.9/8 + 1/9 + 1/10$

Different forms of personalized service systems can be abstracted into a general architecture. That is, the system first collects user information, then models users according to the user information, and then provides personalized service policies and service contents on the basis of the constructed user model. The general architecture of the personalized service system is shown in Figure 2.

In the architecture of the personalized service system, the user information collection module is the basic module of the personalized service system. Since the personalized service is customized for the user, no matter what kind of personalized service it is, the collection of user information is the basis of the personalized service.

According to the analysis of the personalized service architecture, combined with the structural characteristics of ERDDS (one is that the resource description information is stored on the proxy server; the other is that the user's task request processing is implemented on the scheduling

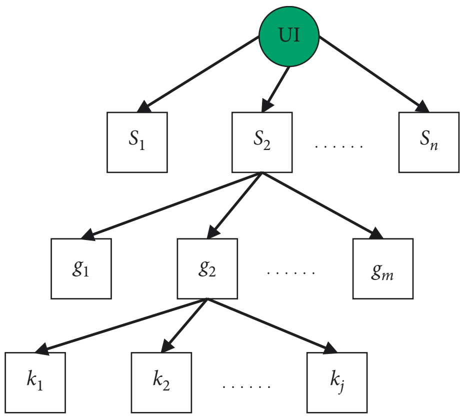


Figure 1: User interest tree.

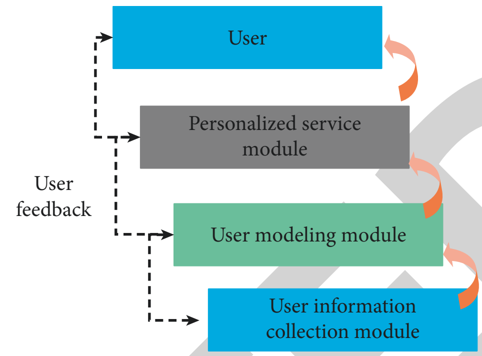


Figure 2: General architecture of personalization services.

server), the personalized service system of this system adopts the structure of the combination of client and proxy server, as shown in Figure 3.

The personalized service connotation of this system includes two aspects. One is to select personalized initial resources for new users of the information platform based on the basic information of resource users. The second is to recommend updated resources of interest to the old users of the information platform based on their interest.

The personalized initial resource recommendation process for new users is shown in Figure 4.

The service process of personalized update resource recommendation service for old users is shown in Figure 5.

In order to accurately express the resource distribution rules based on the basic information of users, the weighted uncertainty representation with credibility is used to express them in the following form [20]:

$$R\colon \text{ if } E_i(\omega_i), \text{then } H(\mathrm{CF}(H, E), \lambda). \tag{4}$$

Among them, E_i is the precondition of the rule, which can be either a simple condition or a combined condition formed by connecting multiple simple conditions with AND. H is the conclusion, which can be a single conclusion or a combined conclusion formed by connecting with AND. $\mathrm{CF}(H, E)$ is the credibility of the rule, called the credibility factor (certainty factor) or rule strength. Credibility is a quantitative representation of the degree of belief that something is true, and its initial value is determined by domain experts. λ is the threshold, which sets a limit on the applicability of the corresponding rule, and only when the credibility $\mathrm{CF}(\mathrm{E_i})$ of the precondition $\mathrm{E_i}$ reaches or exceeds this limit, that is, $\mathrm{CF}(E_i) \geq \lambda$, the corresponding rule is likely to be applied. $\omega_i\,(i = 1, 2, \ldots, n)$ is a weighting factor whose values are all given by domain experts:

$$\begin{gathered}\sum_{i=1}^{n} \omega_i = 1\{0 \leq \omega_i \leq 1, \quad (i = 1, 2, \ldots, n)\} \\ R = \begin{cases} \text{application rules}, & \text{if } \mathrm{CF}(E_i) \geq \lambda, \\ \text{does not apply rules}, & \text{otherwise}. \end{cases}\end{gathered} \tag{5}$$

In the actual processing, the initial decision tree should be prebranched. The generated initial decision tree can be pruned by the prepruning algorithm and the postpruning algorithm, and a pessimistic estimate is used in the pruning process to compensate for the optimistic bias in tree generation. The algorithm extracts classification rules from the resulting decision tree, creates a rule for each path from the root to the leaf, and forms a rule set. Through decision tree classification, the following distribution rules can be obtained:

$$\begin{aligned}
&R_1\colon \text{ if } E_8 \text{ is } A \text{ and } E_9 \text{ is } A \text{ and } E_{12} \text{ is } A, \text{ then result is } Y\,(CF = 0.8, \lambda = 0.6),\\
&R_2\colon \text{ if } E_1 \text{ is } A \text{ and } E_3 \text{ is } B \text{ and } E_4 \text{ is } C \text{ and } E_5 \text{ is } C \text{ and } E_6 \text{ is } A \text{ and } E_9 \text{ is } A \text{ and } E_{12},\\
&\qquad \text{is } A, \text{ then result is } Y\,(\mathrm{CF} = 0.8, \lambda = 0.5),\\
&R_3\colon \text{ if } E_1 \text{ is } B \text{ and } E_2 \text{ is } B \text{ and } E_3 \text{ is } C \text{ and } E_5 \text{ is } C \text{ and } E_6 \text{ is } A \text{ and } E_9 \text{ is } A \text{ and } E_{11},\\
&\qquad \text{is } A \text{ and } E_{12} \text{ is } B, \text{ then result is } Y\,(\mathrm{CF} = 0.7, \lambda = 0.7),\\
&R_4\colon \text{ if } E_3 \text{ is } B \text{ and } E_4 \text{ is } B \text{ and } E_6 \text{ is } C \text{ and } E_7 \text{ is } C \text{ and } E_8 \text{ is } A \text{ and } E_9 \text{ is } A \text{ and } E_{12},\\
&\qquad \text{is } A, \text{ then result is } Y\,(\mathrm{CF} = 0.9, \lambda = 0.6),\\
&R_5\colon \text{ if } E_7 \text{ is } C \text{ and } E_9 \text{ is } C \text{ and } E_{12} \text{ is } C, \text{ then result is } Y\,(CF = 0.8, \lambda = 0.6).
\end{aligned} \tag{6}$$

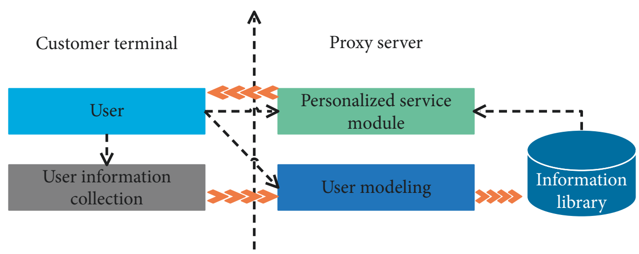


Figure 3: Personalized service architecture.

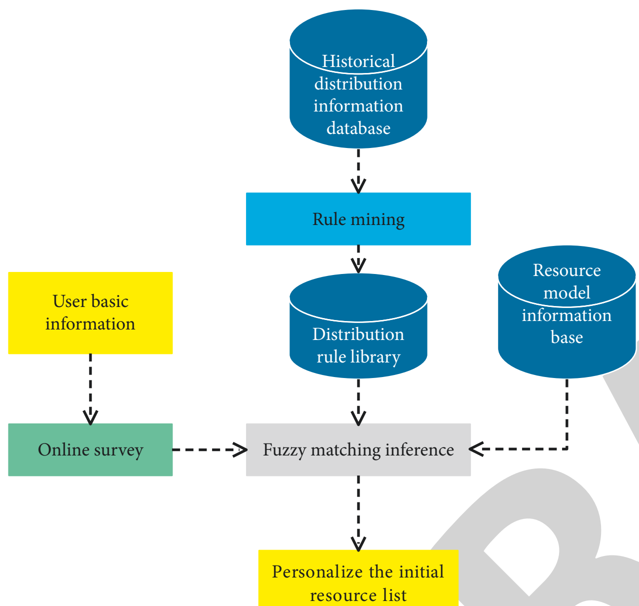


Figure 4: Flow chart of personalized initial resource recommendation for new users.

In the selection of distribution rules, a goal-guided reverse reasoning is used, and fuzzy matching and fuzzy reasoning are used to realize the selection of distribution rules. According to the basic information of the user, fuzzy matching is used to select the delivery rule for the user. If there is no matching rule, fuzzy reasoning can be used to generate new distribution rules to supplement the distribution rule base and continuously enrich the distribution rule base.

Assumptions:

Set of rules: $S_R = \{R_1, R_2, \ldots, R_i\}, \quad i \in N$

Rule: $R_i = \{S_E, S_C, \mathrm{CF}, \lambda, \mathrm{Grade}, \mathrm{Subject}\}$. Among them, the set of preconditions is $S_E = \{E_1(\varpi_1) \wedge E_2(\varpi_2) \wedge \cdots \wedge E_1(\varpi_1)\}$, the set of conclusions is $S_C = \{C_1 \wedge C_2 \wedge \cdots \wedge C_1\}$, CF is the credibility, λ is the rule heuristic threshold, Grade represents the type of school, and Subject represents the characteristic theme of the school.

Known: the set of preconditions $S_E' = \{E_1' \wedge E_2' \wedge \cdots \wedge E_1'\}$, Grade = grade, and Subject = subject.

The fuzzy matching process is as follows:

(1) First, the algorithm extracts a set of rules that satisfy Grade = grade and Subject = subject from the rule set S_R to form the initial rule set S_{R1}.

(2) According to the evidence set S_E', the algorithm selects the rule set that meets the evidence and the preconditions from the initial rule set S_{R1} to form the intermediate rule set S_{R2}.

(3) The algorithm selects a rule R_i from the intermediate rule set S_{R2} to obtain the precondition set S_{Ei} and the threshold λ_i and calculates the total matching degree $\delta_i(S_{Ei}, S_E')$ of the condition and the evidence. If $\delta_i(S_{Ei}, S_E') < \lambda_i$, the algorithm selects the next rule R_{i+1} from the intermediate rule set S_{R2} to recalculate; otherwise, the rule R_i is stored in the target rule set S. The algorithm loops through each rule in S_{R2} until the end goes to (4).

(4) After conflict resolution processing, the algorithm obtains effective rules from the target rule set S, such as selecting the rule R_i with the largest matching degree. If the need is met, the rule is quoted and the end returns. If it is not satisfied, the algorithm tries fuzzy reasoning.

The fuzzy reasoning process is as follows:

(1) First, the algorithm extracts a set of rules that satisfy Grade = grade and Subject = subject from the rule set S_R to form the initial rule set S_{R1}.

(2) According to the evidence set S_E', the algorithm selects the rule set that meets the evidence and the preconditions from the initial rule set S_{R1} to form the intermediate rule set S_{R2}.

(3) The algorithm selects a rule R_i from the intermediate rule set S_{R2} to obtain the precondition set S_{Ei} and conclusion set S_{Ci}, credibility of the rules CF_i, and the threshold λ_i and calculates the total matching degree $\delta_i(S_{Ei}, S_E')$ of the condition and the evidence. If $\delta_i(S_{Ei}, S_E') < \lambda_i$, the algorithm selects the next rule R_{i+1} from the intermediate rule set S_{R2} to recalculate. Conversely, the algorithm computes the intersection $\wedge E_i$ of the combined conditions in the condition set S_{Ei}.

(4) The algorithm constructs the fuzzy relation $R_i(\wedge E_i, S_C)$ between the combination condition intersection $\wedge E_i$ and the conclusion set S_C using a fuzzy relation method such as Zade method (R_a or R_m) or Memdney method (R_S).

(5) The algorithm finds the intersection $\wedge E'$ of the combined evidence.

(6) According to the fuzzy hypothesis reasoning, the algorithm obtains the conclusion fuzzy set C_i' by synthesizing the intersection $\wedge E'$ and $R_i(\wedge E_i, S_C)$ of the evidence.

(7) Through the total matching degree $\delta_1(S_{Ei}, S_E')$ of conditions and evidence, the credibility $CF_{E'}$ of evidence, the credibility CF_i of rules, and the credibility CF_i' of the theoretical fuzzy set C_i' are calculated by the algorithm.

(8) The algorithm jumps to (3) and loops through each rule in S_{R2} until the end, after which the algorithm goes to (9).

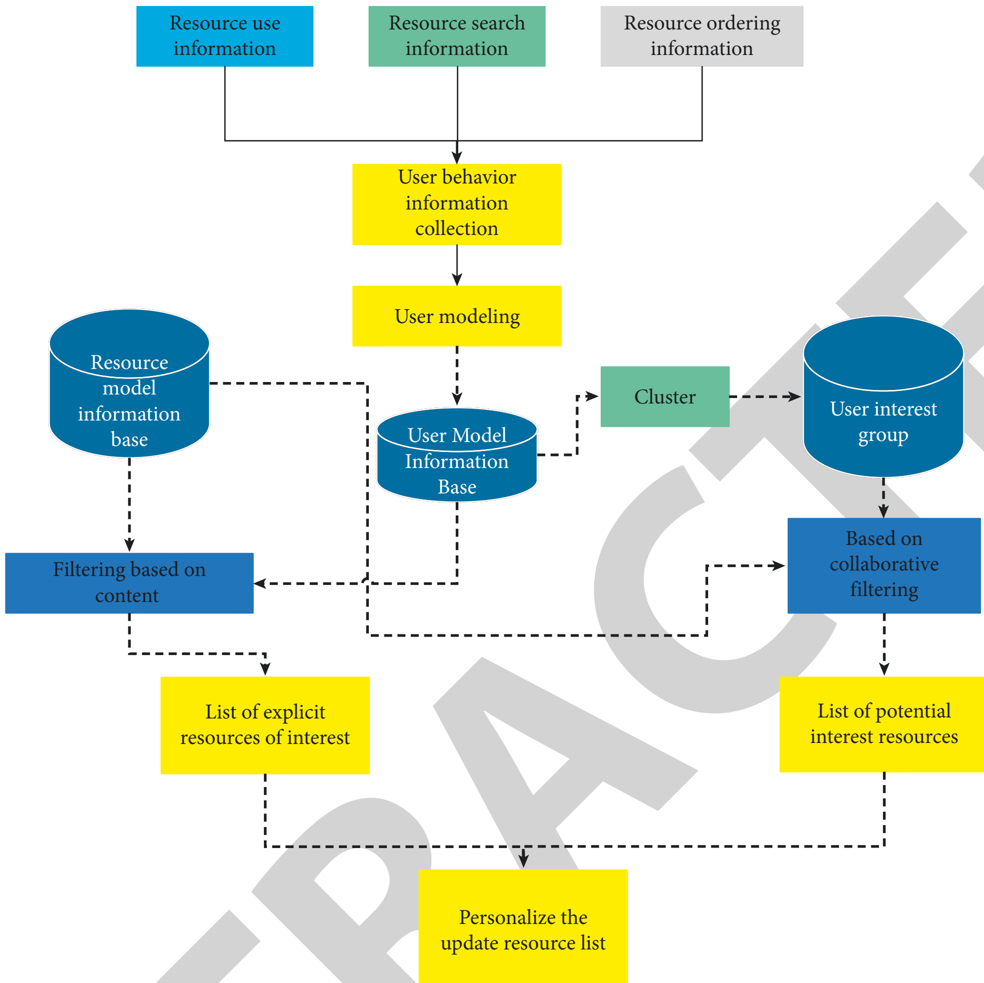


Figure 5: Flow chart of personalized update resource recommendation service.

(9) After conflict resolution processing, the algorithm selects effective rules, such as the conclusion C_i'° with the greatest reliability. If the need is met, the algorithm cites the conclusion C_i' and stores the deduced rules in the distribution rule base and returns at the end.

For vector space models, the commonly used methods are Euclidean distance, cosine distance, and inner product.

For any two vectors $X = (x_1, x_2, \ldots, x_n)$ and $X' = (x_1', x_2', \ldots, x_n')$:

The Euclidean distance is

$$d(X, X') = \left(\sum_{i=1}^{n} (x_i - x_i)^2\right)^{1/2}. \quad (7)$$

The cosine distance is

$$d(X, X') = \frac{\sum_{i=1}^{n} x_i x_i'}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} x_i'^2}}. \quad (8)$$

The inner product is

$$d(X, X') = \sum_{i=1}^{n} x_i x_i'. \quad (9)$$

The greater the distance between the user interest keyword vector and the resource description keyword vector, the greater the similarity between them, and vice versa.

The first step of data mining is data preparation, which includes data selection, data preprocessing, and data transformation. For the clustering of user interest groups, it corresponds to clustering keyword selection, user interest model standardization, and interest degree conversion.

3.1. Interest Degree Conversion. Interest is the degree to which people are interested in an event. For the keywords in the user interest model, the larger the weight (the more occurrences), the more the user is interested in the keyword. Therefore, the weight of keywords is used here to represent the user's interest. Since the degree of interest is set to be between [0, 1], the range transformation is used to normalize the weight of each keyword, and then the degree of interest is expressed.

$$I\ D\left(k_j\right) = \frac{\omega\left(k_j\right) - \mathrm{Min}\,(\omega\,(K))}{\mathrm{Max}\,(\omega\,(K)) - \mathrm{Min}\,(\omega\,(K))}, \quad k_j \in K. \tag{10}$$

In the formula, $I\ D(k_j)$ represents the user's interest in the keyword k_j; $\omega(k_i)$ represents the weight of the keyword k_j; $\mathrm{Min}\,(\omega\,(K))$ represents the minimum value of all keyword weights in the user interest model; and $\mathrm{Max}\,(\omega\,(K))$ represents the maximum value of all keyword weights in the user interest model.

3.2. Clustering Keyword Selection. Although the keywords of the user interest model can reflect the user's interest topics, it is inevitable that there will be some keywords that the user is not very interested or even not interested in. If these keywords that users are not very interested in are also used for interest clustering, it will inevitably increase the calculation amount of the interest clustering algorithm and also affect the quality of the clustering. Therefore, it is necessary to select the keywords participating in the interest clustering, and only those keywords with a high degree of user interest can be used for the interest clustering. Therefore, before the user interest group clustering, the keywords smaller than the initial interest degree threshold Min_ID are removed, and the keywords with higher interest degree are selected for clustering, and the initial interest degree threshold Min_ID is determined by the user.

3.3. Standardization of User Interest Model. The interests of users are different, the number of keywords involved in the corresponding interest model is different, and the degree of interest in a certain aspect is also different. In order to compare the similarity of user interest models, the user interest models need to be normalized. The standardized processing method is to perform a union operation on all user interest keywords with the same knowledge background (same subject and school stage) as a reference model of all user interest keyword vectors. For the keyword interest degree, the keywords belonging to the user themselves, the interest degree remain unchanged. However, if it does not belong to the user, the interest degree is set to 0.

For example, the set of users with the same knowledge background is $U = \{u_1, u_2, \ldots u_n\}$, and the interest keyword vector of each user u_i is $K_{ui} = \left\{k_1, k_2, \ldots, k_{m_{mi}}\right\}$, then the reference model of all user interest keyword vectors is

$$\begin{aligned} K' &= K_{u1} \cup K_{u2} \cup \cdots \cup K_{un} \\ &= \{k_1', k_2', \ldots, k_m'\}. \end{aligned} \tag{11}$$

All user interest keyword vectors take K' as the reference model and are correspondingly transformed into $K_{ui}' = \{k_1', k_2', \ldots, k_m'\}$. The interest degree of each keyword k_J' of K_{ui}' is processed as follows:

$$I\ D\left(k_j'\right) = \begin{cases} I\ D\left(k_j\right), & k', \in K_{ui}, \\ 0, & \text{otherwise}. \end{cases} \tag{12}$$

After obtaining user interest groups, the keyword entries in each interest group model are consistent with the clustered user keyword vector reference model. The interest degree of a keyword is the average value of the corresponding keyword interest degree in all user interest models in the group, namely,

$$\mathrm{ID}\left(k_i\right) = \frac{\sum_{j=1}^{m} ID\left(k_i\right)_j}{m}. \tag{13}$$

In the formula, $\mathrm{ID}\,(k_i)_j$ represents the degree of interest of user u_j to the keyword k_i and m represents the number of users in the interest group.

For personalized recommendation systems, the main recall rate and precision rate are generally used to evaluate the performance of the system. For a good recommender system, its precision rate and recall rate should be as large as possible. However, it is often not possible to have the best of both worlds. Sometimes the recall rate is high, the precision rate is low, and when the precision rate is high, the recall rate is low.

(1) Recall Rate. The recall rate refers to whether the recommended amount of resources can cover all qualified resource records and is used to indicate the ability of the system to recommend related resources. It can be expressed as the total number of recommended related resources divided by the total number of related resources in the resource upgrade package.

$$\text{recall rate} = \frac{\text{recommended relevant resources}}{\text{total relevant resources in the system}} \times 100\%. \tag{14}$$

(2) Precision Rate. The precision rate is the degree to which the recommended resources meet the recommended purpose and the ability to reject irrelevant resources. It can be expressed as the total amount of recommended related resources divided by the total amount of recommended resources.

$$\text{precision rate} = \frac{\text{recommended relevant resources}}{\text{total recommended resources}} \times 100\%. \tag{15}$$

4. Application of Improved Machine Learning and Fuzzy Algorithm in Educational Information Technology

The corresponding objects of this system are high school learners, teachers, school administrators, and parents of students. The system will be divided into four modules: adaptive module, student portrait module, early warning and intervention module, and learning incentive module. Figure 6 shows an educational information system based on improved machine learning and fuzzy algorithms.

As shown in Figure 7(a), the distance education intelligent decision support system is mainly composed of the following components, namely, database management system, model base management system, knowledge base management system, and decision information output management. Among them, the database includes courseware multimedia database, student information database,

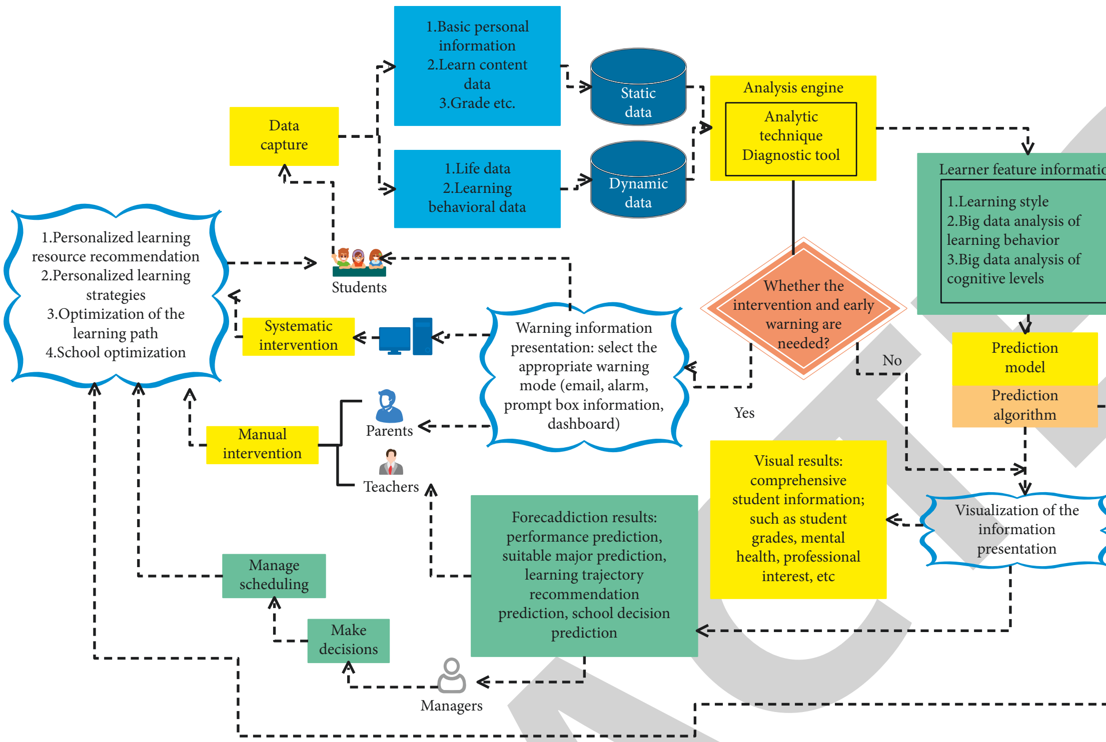
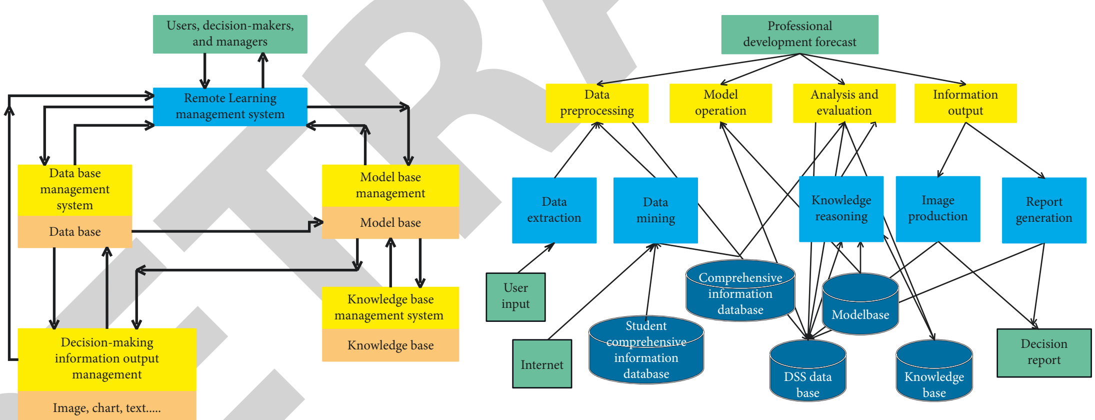


Figure 6: Educational information system based on improved machine learning and fuzzy algorithms.



(a) (b)

Figure 7: System function modules. (a) Logical structure of the intelligent decision support system. (b) Logical structure of the professional forecasting module.

and comprehensive information database. The database management system is responsible for the management and daily maintenance of each database. The model library stores various models required for decision-making. The model library management system is responsible for classifying and maintaining models, and supporting model generation, storage, query, operation, and analysis applications. The knowledge base stores all kinds of knowledge, and the knowledge base management system is responsible for knowledge reasoning, machine learning, and knowledge maintenance (including addition, deletion, and modification). The remote teaching management system completes the man-machine dialogue. In addition, decision information output management is responsible for displaying decision information in the form of charts, images, text, reports, etc.

As shown in Figure 7(b), professional development prediction consists of four modules, namely, data preprocessing, model operation, analysis and evaluation, and information output. First, the user inputs requirements through the distance education management system, and the data extraction module extracts the user input and the data mined by the data mining module is sent to the data preprocessing module for data preprocessing and then stored in the ADSS database. At the same time, the processed data is sent to the model operation module. Next, the model operation module calls the corresponding model in the model

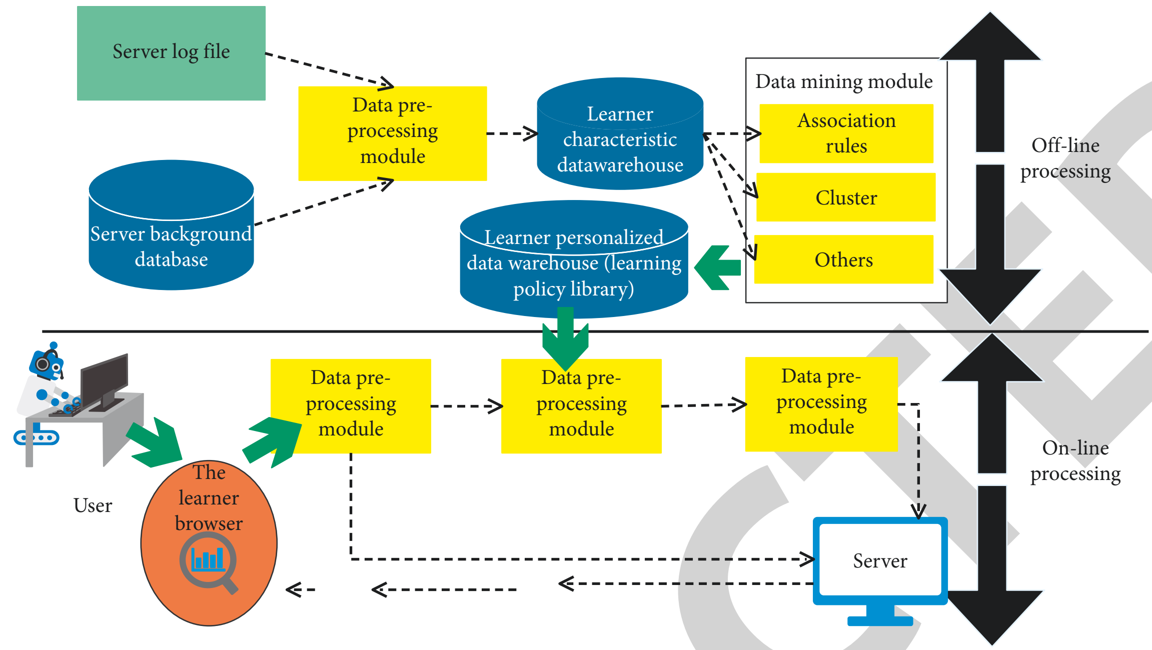


Figure 8: Web-based personalized learning system.

Table 1: Statistical table of the processing effect of educational information technology.

Number	Information processing	Number	Information processing	Number	Information processing
1	91.75	22	91.15	43	89.20
2	91.99	23	89.52	44	91.44
3	93.66	24	89.25	45	89.98
4	90.30	25	94.72	46	90.80
5	89.44	26	91.04	47	93.31
6	89.74	27	92.73	48	91.31
7	90.37	28	89.59	49	90.44
8	94.76	29	92.95	50	94.08
9	90.59	30	89.58	51	94.67
10	92.08	31	93.45	52	93.66
11	94.82	32	90.31	53	94.15
12	92.67	33	89.39	54	93.84
13	91.91	34	89.76	55	91.85
14	89.80	35	89.17	56	94.49
15	89.25	36	89.01	57	94.00
16	93.70	37	89.90	58	92.26
17	89.39	38	89.30	59	94.01
18	91.43	39	92.16	60	90.15
19	93.53	40	92.53	61	91.26
20	90.04	41	90.79	62	94.66
21	90.98	42	90.27	63	89.20

library and uses the corresponding prediction algorithm to operate on the processed data to obtain multiple different prediction results. However, the result cannot be directly output as auxiliary decision-making information, and the occurrence probability, reliability, and error of various prediction results must be evaluated through knowledge reasoning analysis. This process mainly uses models, knowledge, cases, experiences, and other data in the comprehensive information database and DSS database to reason, analyze, evaluate, and store the results approved by decision makers as a kind of knowledge in the knowledge base as an auxiliary decision-making basis. Finally, an auxiliary decision-making report is formed through the information output module.

Figure 8 is the main architecture of the web-based personalized learning system in terms of technology.

This paper verifies the effect of the educational information technology model constructed. First, the processing effect of the system model in this paper on educational information is tested. The statistical test results are shown in Table 1.

It can be seen from the above research that the system proposed in this paper can play an important role in the processing of educational information technology.

Table 2: Teaching evaluation effect.

Number	Teaching evaluation	Number	Teaching evaluation	Number	Teaching evaluation
1	81.25	22	78.18	43	74.21
2	74.09	23	77.33	44	76.03
3	75.61	24	85.51	45	82.61
4	81.50	25	77.27	46	85.19
5	78.63	26	76.11	47	82.71
6	75.38	27	80.80	48	85.65
7	81.78	28	82.41	49	83.44
8	78.80	29	84.13	50	76.30
9	74.78	30	76.20	51	80.57
10	81.96	31	78.95	52	78.60
11	77.12	32	85.29	53	80.43
12	83.45	33	85.75	54	74.78
13	76.94	34	74.59	55	76.77
14	79.27	35	78.89	56	74.04
15	84.31	36	82.93	57	79.47
16	82.41	37	77.43	58	78.15
17	76.62	38	80.97	59	80.73
18	79.04	39	80.88	60	84.52
19	77.86	40	77.81	61	74.34
20	74.95	41	76.91	62	78.70
21	77.83	42	82.38	63	85.30

Through the above experimental research (as shown in Table 2), we can see that the educational information system based on the improved machine learning and fuzzy algorithm proposed in this paper has a good effect.

5. Conclusion

Due to the rapid development of modern science and technology, people who are engaged in specific technical work will fall behind if they do not capture the latest information in time. Therefore, in the application of educational information technology, it is necessary to strengthen information awareness, improve information quality, and pay special attention to educational information and service objects. People who are engaged in technical work must have excellent skills without making themselves "technical idealists." The corresponding objects of this system are high school learners, teachers, school administrators, and parents of students. Moreover, the system will be divided into four modules: adaptive module, student portrait module, early warning intervention module, and learning incentive module. The experimental research shows that the educational information system based on improved machine learning and fuzzy algorithm proposed in this paper has a good effect.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Disclosure

Due to the change of team members, Hetiao Hong left to work in other departments, so the author Hetiao Hong was removed, and Zhangfu Wang who contributed to this work was added.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was sponsored by Jing Hengyi School of Education, Hangzhou Normal University.

References

[1] S. F. M. Alfalah, "Perceptions toward adopting virtual reality as a teaching aid in information technology," *Education and Information Technologies*, vol. 23, no. 6, pp. 2633–2653, 2018.

[2] G. Cooper, H. Park, Z. Nasr, L. P. Thong, and R. Johnson, "Using virtual reality in the classroom: preservice teachers' perceptions of its use as a teaching and learning tool," *Educational Media International*, vol. 56, no. 1, pp. 1–13, 2019.

[3] J. Zhao, X. Xu, H. Jiang, and Y. Ding, "The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies," *BMC Medical Education*, vol. 20, no. 1, pp. 127–210, 2020.

[4] S. J. Bennie, K. E. Ranaghan, H. Deeks et al., "Teaching enzyme catalysis using interactive molecular dynamics in virtual reality," *Journal of Chemical Education*, vol. 96, no. 11, pp. 2488–2496, 2019.

[5] S. F. M. Alfalah, J. F. M. Falah, T. Alfalah, M. Elfalah, N. Muhaidat, and O. Falah, "A comparative study between a virtual reality heart anatomy system and traditional medical teaching modalities," *Virtual Reality*, vol. 23, no. 3, pp. 229–234, 2019.

[6] M. Reymus, A. Liebermann, and C. Diegritz, "Virtual reality: an effective tool for teaching root canal anatomy to undergraduate dental students–a preliminary study," *International Endodontic Journal*, vol. 53, no. 11, pp. 1581–1587, 2020.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9798586, 1 page
https://doi.org/10.1155/2023/9798586

Retraction

Retracted: Exploring the Influence of Big Data Technology on the Innovation of the Enterprise Economic Management Mode

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wu, "Exploring the Influence of Big Data Technology on the Innovation of the Enterprise Economic Management Mode," *Security and Communication Networks*, vol. 2022, Article ID 6241182, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6241182, 13 pages
https://doi.org/10.1155/2022/6241182

Research Article

Exploring the Influence of Big Data Technology on the Innovation of the Enterprise Economic Management Mode

Yingning Wu

College of Economics and Management, Hefei University, Hefei, Anhui 230601, China

Correspondence should be addressed to Yingning Wu; wyn@hfuu.edu.cn

Received 9 February 2022; Revised 1 April 2022; Accepted 18 April 2022; Published 19 May 2022

Academic Editor: Chin-Ling Chen

The aim of this study is to help financial enterprises establish a solid foundation in a big data ecosystem (BDE) and fully play their competitive edges in the fierce business competition. This paper innovatively conducts research studies on the reform of enterprise financial management. First, it expounds on the relevant concepts of big data technology (BDT) alongside financial management. Afterward, the impact of BDT on enterprise financial management is examined. Consequently, a financial enterprise-oriented economic management model is proposed by integrating BDT, enterprise economic management (EEM), and enterprise performance evaluation methods. Further, an experiment is designed to verify the proposed model on estimating the annual enterprise operating income (OI) and human resource (HR) structure. The research findings show that the predicted and actual OI error is within 6%. The predicted enterprise personnel deployment only deviates slightly from actual personnel deployment. Thus, the proposed economic management model can accurately predict OI and cost. It provides a reliable reference for enterprise innovation management, optimizing recruitment conditions, sales strategy, and development strategy. Lastly, the research deficiencies are pointed out alongside corresponding suggestions.

1. Introduction

The 21st century is the era of network science and technology (S&T), when computer and networking technology (CNT) is undergoing a new revolution, thus bringing dramatic social changes [1]. New technologies are seeing broader applications. For example, mobile computing (MC), Internet of Things (IoT), and edge computing (EC) technologies [2] have expanded to social media and system manufacturing. It replaces the traditional communication approaches, leading humanity into the big data era [3]. In China, the dawn of the big data era can trace back to the action plan for promoting the development of big data in 2015 [4]. Various technical exchange platforms were launched to offer enterprises the latest industrial updates and encourage industrial innovations. In financial sectors, economic management safeguards financial and monetary transactions and thus is indispensable for the development of financial enterprises [5].

However, with a deeply rooted traditional management concept, many enterprises are still skeptical about data sharing services (DSS), such as data interactive management [6]. Big data expands the scope of knowledge. Thus, today's enterprise management (EMA) involves more than just business experience and management concepts but also real-time transaction data and online resources. At the same time, EMA [7] must be innovated by integrating various CNT into the big data ecosystem (BDE) to survive and prosper with a solid technical foundation in international competition [8]. Enterprise competition has a long history and exists throughout an entire enterprise life cycle with a legal appearance. Fajar believed that the competition law must be designed pragmatically to keep up with rapidly changing business models. Generally, it was agreed upon by business actors to create a more dynamic market environment and more prosperous consumers [9]. Hence, enterprise competition management is necessary, and most EMA models are affected by the mainstream setting. Shao et al. analyzed environmental regulation (ER) on economic sustainable development (SD), enterprise innovation on ER, and enterprise initiative on ER from the perspective of the enterprise ecosystem. They also examined the relationship between ER and enterprise innovation and social security, as well as the integration of eliminated enterprises [10]. Lin et al. held that maintaining competitive advantage and improving innovation performance in the BDE was crucial for enterprise development. They found that the positive role of management power and network centrality was more significant in the BDE. These findings enriched the research on high-tech enterprise innovation from the perspective of portfolio governance and contributed to the literature research on enterprise innovation in the BDE [11]. Thus, network technologies, such as BDT, can improve the EMA significantly.

After a literature review, this paper proposes a new management and innovation concept for financial enterprises using BDT based on the existing economic management model. First, the relevant concepts of big data are explained alongside the economic management mode and influencing factors. Then, an innovative integrated economic management model is proposed. The model performance is verified by two experiments. One is to estimate the operating income (OI) and compare the result with the actual OI. The other is to estimate the enterprise's human resource (HR) structure. The results provide some opinions for the innovation of enterprise economic management (EEM). The research innovation lies in evaluating the rationality of the proposed model by comparing the historical data with the model prediction data.

2. Methods

2.1. Big Data Concept. Big data is primarily used in innovative modes of high-volume, high-velocity, and high-variety information assets [12]. Big data can be used in decision-making (DM), information mining, data analysis, and service recommendations. For example, by analyzing user's preference big data, businesses can recommend products and services to users as a new "marketing" approach [13]. Anyway, high volume is the most revealing but not the only feature of big data. Table 1 shows the presentation characteristics of big data [14].

The BDE has significantly changed the world by integrating traditional single-data networks. Metaphorically, today's world is a vast digital network surrounded by a BDE. Hence, users can obtain information concerning almost all social spheres through network terminals. Figure 1 summarizes the characteristics of BDT [15].

Table 2 explains the specific meanings of the four characteristics of BDT.

The BDE has brought a different impact on enterprises. In BDE, people can extract valuable data to understand and analyze things more clearly because data correlations are much stronger. Figure 2 lists people's mind set transformation from four aspects in BDE.

2.2. Enterprise Economic Management. An enterprise is a socioeconomic unit with a unique structure that focuses on economic exchanges [16]. The traditional economic concept divides the enterprises' business activities into two types [17], as outlined in Figure 3.

As illuminated in Figure 3, EMA includes business and production management and economic management. The former stresses commodities operations, while the latter focuses on businesses values.

Traditionally, economic management concerns only enterprises' basic business activities, such as goods collection, investment amount, fund use, and profit distribution [18].

The emerging economic management considers the complex enterprise architecture instead of the enterprise itself and pays more attention to the balance of funds and cost. The new economic management model factors cost control and performance evaluation, most of which are a series of standard processes for monitoring and assessing the enterprise economic activities. Different enterprises have different management objects [19]. Figure 4 depicts the several characteristics of modern EEM.

2.3. Impact on Enterprise Economic Management. The BDE and economic globalization both have a far-reaching impact on the management and operation of enterprises and people's social behaviors. Economic management is the cornerstone of an enterprise. Thus, the economic management concepts, environment, and technologies must all be innovated for enterprises to adapt to the new era [20].

Particularly in BDE, the offline marketing environment has transformed into virtual, online, and digital modes, from face-to-face communication to networked exchange with economic objects. Moreover, the enterprise can obtain instant real-time feedback from local markets. Figure 5 specifies the economic regulation from several aspects.

2.4. Economic Management Model. This section strives to propose an optimized economic management model. Generally, the EEM level can be evaluated through the internationally recognized enterprise operation indexes to improve EMA. Accordingly, the proposed economic management model selects the EEM ability as the benchmark and divides it into five levels: the initial level, the basic level, the specification level, the management level, and the optimization level. Meanwhile, the five levels will be gradually upgraded to the subsequent higher levels through adaptation to the previous levels, as illustrated in Figure 6.

As illustrated in Figure 6, enterprises can self-evaluate, identify, analyze, and repair loopholes through the abovementioned five levels using technical innovation to balance reform and sustainable development [21]. Table 3 enumerates the five EEM levels with their meaning.

2.5. Performance Evaluation System (PES). According to the literature review, humans always weigh the input-output Ratio (IOR) during production activities for life, survival, or reproduction. Some scholars believe IOR to be the early statement of performance evaluation. Therefore, the idea of

Table 1: Big data characteristics.

Characteristic name	Characteristic content
Huge amount of data	Numerous network terminals can record many data, such as user trajectory and browsing data. Each user can be provided with a gigantic data pool
Diverse data types	Various sensors, social platforms, and network terminals generate semistructured and unstructured data, making the data types complex and diverse
Efficient processing rate	By analyzing big data, businesses can push similar products in real-time according to users' browsing interests, such as the well-known amazon and JD shopping platforms
Data authenticity	With their actual content partially ignored, big data provide a reference basis for decision-makers to extract valuable information from the real data center (DC), which is the basis for obtaining knowledge

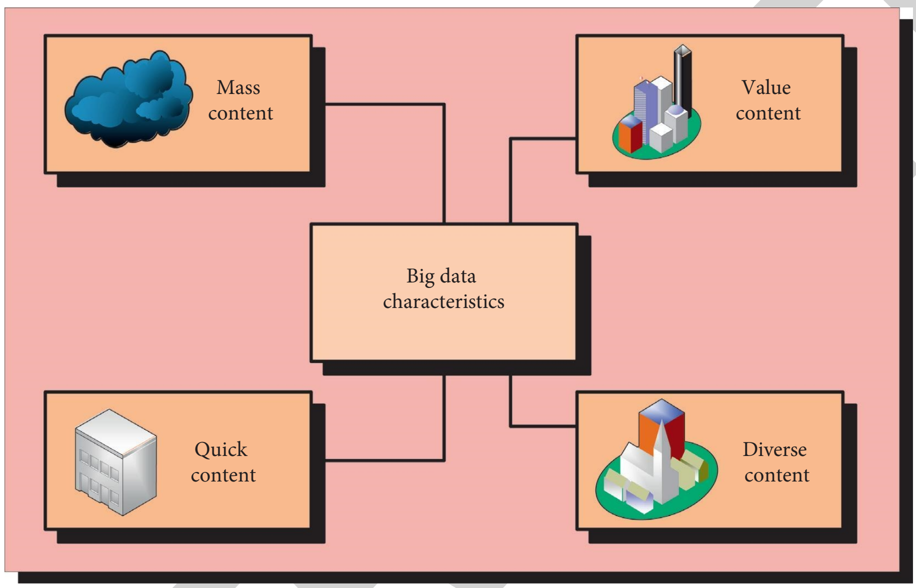


Figure 1: Characteristics of BDT.

Table 2: Connotation of BDT characteristics.

Characteristic	Connotation
Massive content	BDT uses a tremendous amount of data volume
Diverse content	BDT involves various data types from all social levels
Quick content	BDT utilizes CNT to achieve fast data transmission
Value content	BDT analyzes large amounts of low-density data

performance evaluation might have existed as early as a primitive society. The current performance evaluation theory [22] mainly includes the following:

2.5.1. Asset-Based Valuation Method. It includes cost replacement, market value, and accounting value. Of these, cost replacement is the final enterprise cost after mergers and acquisitions, removal of related consumption, and devaluation. Market value is an effectiveness hypothesis used in macro conditions according to the market price fluctuation. The accounting value is determined according to the asset amount of the traditional accounting book, which is a static valuation method and generally expressed in the balance sheet. Figure 7 displays the specific structure.

2.5.2. Relative Valuation Method. The relative valuation indexes can intuitively display the enterprise's worth. For example, the price/earnings ratio (P/E) directly relates to the enterprise's profitability and operation effect. The market sales rate is a key standard for enterprise listing and development. These indexes are used either for the enterprise loss evaluation with strict requirements or investor's fund evaluation. Figure 8 demonstrates the specific structure.

2.5.3. Discounted Cash Flow (DCF) Method. It is a mainstream valuation method focusing on the discounted cash flow (CF) in the enterprise life cycle. In actual situations [23], the DCF method depends on the future CF. In the 1980s, research experts defined liquidity as part of CF. CF can be divided into two types according to payment objects.

Enterprise CF is calculated by the following:

$$V_t = \sum_{t=1}^{n} \frac{\mathrm{FCFF}_t}{(1 + WACC)^t}. \tag{1}$$

Share CF is calculated by the following:

$$V_t = \sum_{t=1}^{n} \frac{\mathrm{FCFF}_t}{(1 + K_c)^t}. \tag{2}$$

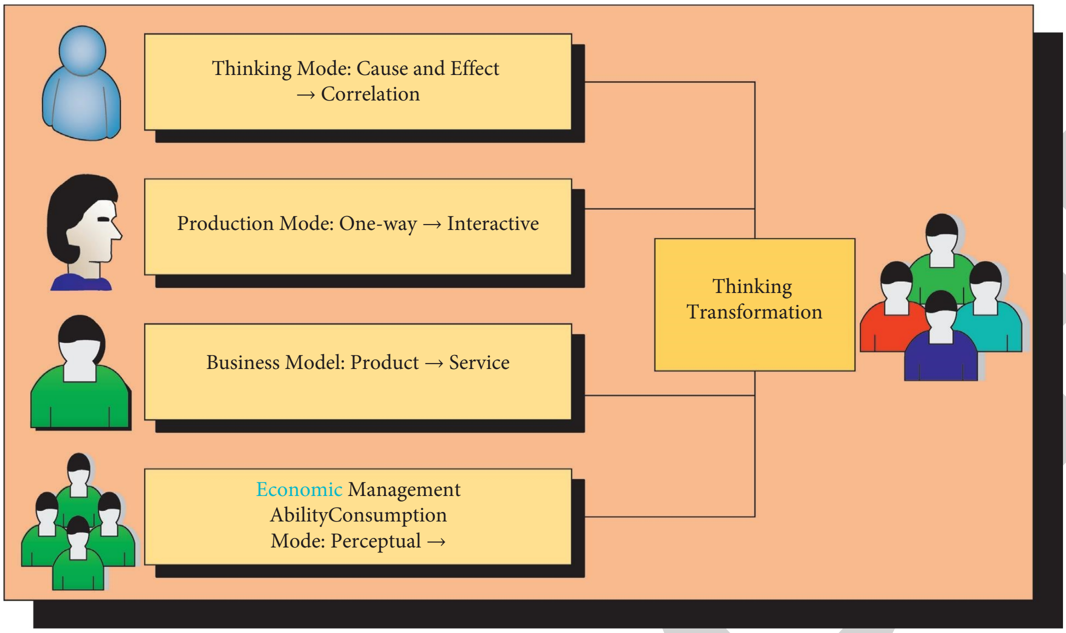


FIGURE 2: Transformation of people's mindset by BDE.

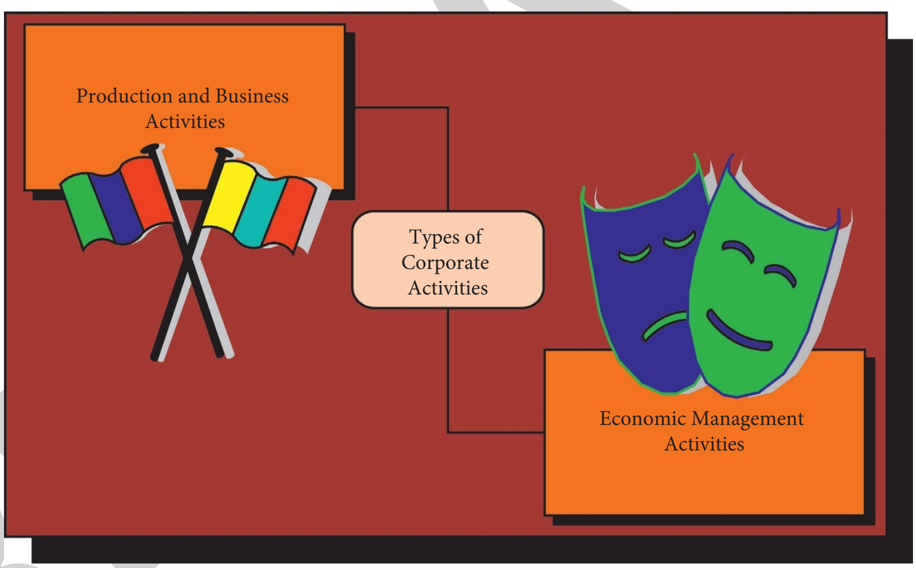


FIGURE 3: Types of business activities.

In (1)–(2), *Vt*, *FCF*, and *WACC* represent value, the free CF, and the weighted average loss, respectively. K_C refers to the cost. *F* means the fixed investment added when sales rise one RMB. *T* stands for the time. (1) and (2) present different cash disbursements.

Debt cost occurs with enterprise financing, namely, weighted average loss. The calculation method is as follows:

$$WACC = K_b \times \frac{B}{V} \times (1 - T) + K_s \times \frac{S}{V}. \quad (3)$$

In (3), K_s, K_b, S, B, and V denote the share cost, pre-tax cost, the share value, the debt, and the total assets of the enterprise.

The enterprise applying the DCF model must comply with the following conditions in terms of the CF criteria. (1) The enterprise discounts according to the CF in the future period. (2) The enterprise must maintain an incoming-fund flow in recent years, and the risks should be controllable.

2.5.4. Option Valuation Method. The accuracy and universality of domestic enterprise performance evaluation have been improved by consulting European and American research [24]. The Black-Scholes (B-S) model was proposed in the 1970s. The conceptual basis is the value of financial options. After optimization, this model is applied in practice

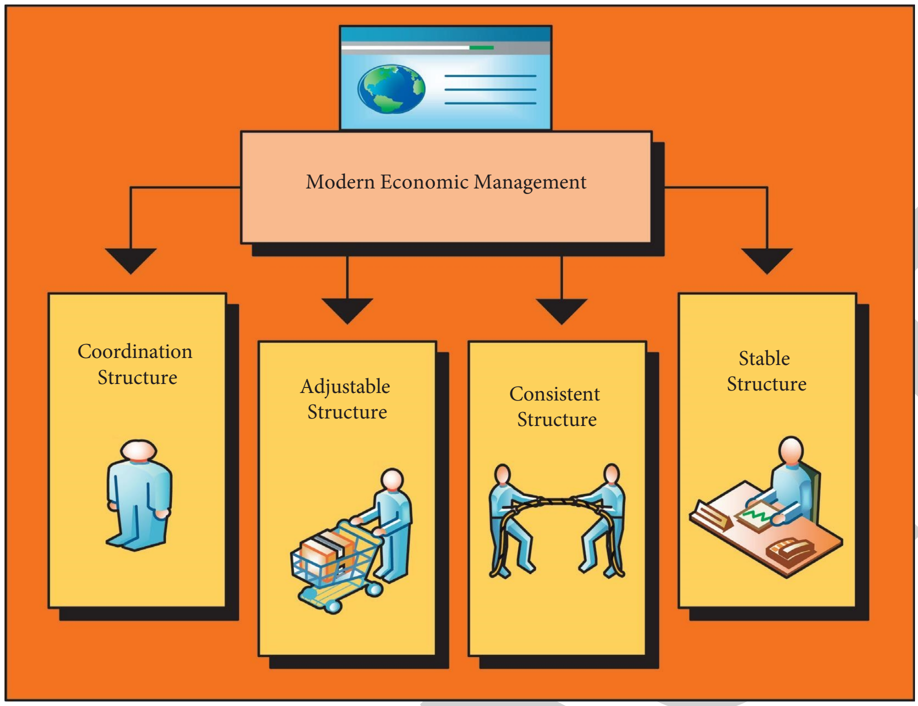


Figure 4: Characteristics of modern EEM.

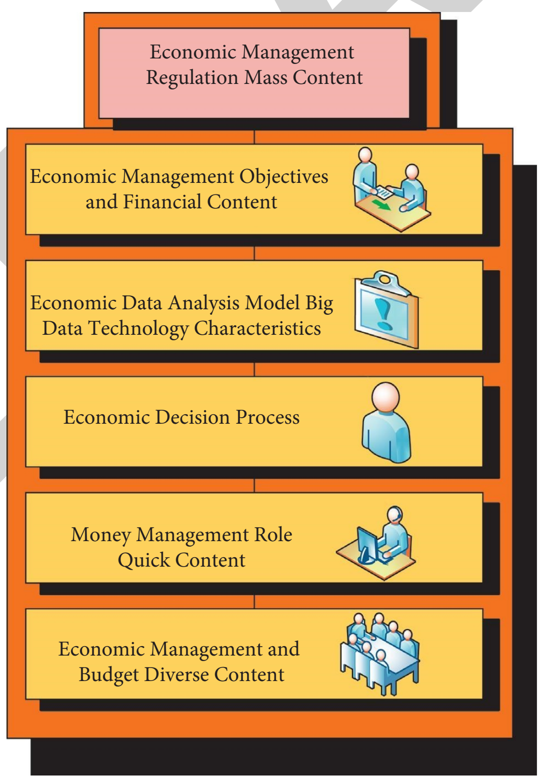


Figure 5: Aspects of EEM regulation.

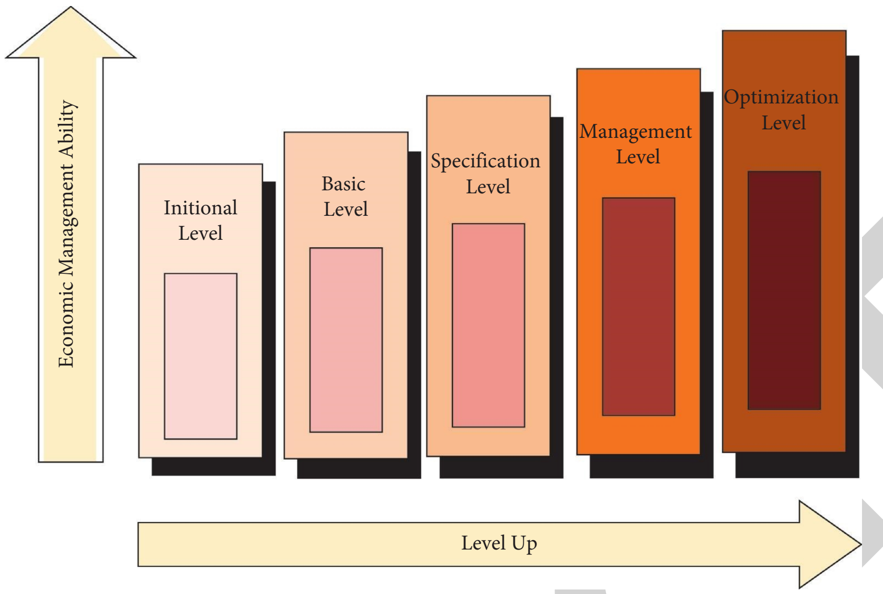


Figure 6: EEM hierarchy chart.

Table 3: Meaning of the EEM level.

Name	Specific meaning
Initial level	At this stage, the enterprise is still in the initial step. Its structure is relatively simple and single. EEM ability is relatively weak. How to optimize the use structure of funds has become the main problem
Basic level	Compared with the initial level, this level emphasizes the cultivation of EEM ability and can have a relatively independent set of institutional strategies. In this link, artificial intervention is common, and the implementation of the fund system lacks effective management
Specification level	At this stage, enterprises begin to transform from a single growth stage to diversification and gradually complete their legal structure. The leadership begins to pay more attention to EEM. The main optimization concern is strengthening the communication between the capital department and other positions
Management level	Enterprises reaching this stage have an integral internal structure and system. They can control and analyze business activities in terms of EMA and provide reference data for performance evaluation. Meanwhile, they can improve their strategic abilities and play a leading role
Optimization level	The enterprise scale is considerable at this stage, and the EMA level will be upgraded to the strategic level. The whole capital flow (CF) system has been perfected, so the financial activities can be changed according to the external environment to guide enterprises to develop an excellent competitive management ability

to explain the actual work of enterprise valuation and expound the methods and principles.

Besides, the B-S model belongs to a binary tree model, and its calculation reads as follows:

$$C_0 = S_0\left[N\left(d_1\right)\right] - Xe^{-r_c t}\left[N\left(d_2\right)\right]. \tag{4}$$

(4) can also be expressed as follows:

$$C_0 = S_0\left[N\left(d_1\right)\right] - PV\left(X\right)\left[N\left(d_2\right)\right]. \tag{5}$$

In (5),

$$d_1 = \frac{\ln\left(S_0 \div X\right) + \left[r_c + \left(\sigma^2 \div 2\right)\right]t}{\sigma\sqrt{t}}. \tag{6}$$

Equ. (6) can also be expressed as follows:

$$d_1 = \frac{\ln\left[S_0/PV\left(X\right)\right]}{\sigma\sqrt{t}} + \frac{\sigma\sqrt{t}}{2}. \tag{7}$$

In (7),

$$d_2 = d_1 - \sigma\sqrt{t}. \tag{8}$$

In (4)–(8), C_0 is the risen value of shares. S_0 denotes the current value of shares. $N\left(d_1\right)$ represents the deviation less than d in the normal distribution. X means the exercise value of options. r_c refers to the compound interest risk-free profit for consecutive periods. t is the maturity time. σ^2 stands for the variance of share return of continuous compound interest.

2.5.5. Economic Value Added (EVA) Evaluation Method. Traditionally, enterprise profit is calculated by debt costs and has limited applications because equity financing will increase enterprise costs [25]. Comparatively, the EVA approach [26] can break through such limitations. A single-stage model is viable to mature enterprises with relatively

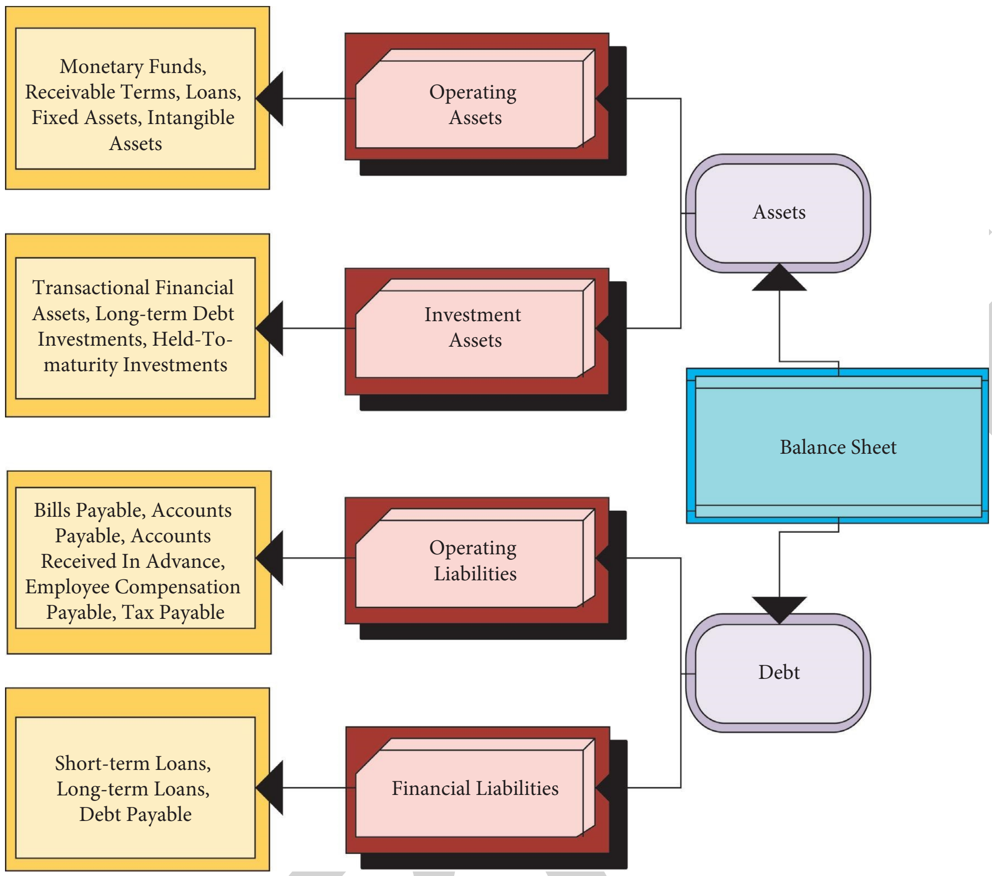


FIGURE 7: Asset valuation structure.

stable organization structures, CF, and development. In such cases, the EVA is recorded as g calculated by

$$V = \frac{IC + EVA}{WACC - g}. \tag{9}$$

EVA refers to the first EVA value in the period of stable growth.

2.6. Enterprise Economic Management Model. Economic management models select different enterprise characteristics according to unique situations. Timely innovating the economic management model will directly impact business strategies. Such transformation is not natural but requires initiative, planning, and structural innovation based on unique enterprise competitive edges. Only a well-built enterprise innovation structure can inject new concepts, proactive elements, and innovation into enterprises. There is a need to build an integrated model to maximize EMA efficacy. Table 4 unfolds the key features of the proposed integrated economic management model from two dimensions.

Subsequently, this section selects several financial enterprises with maladaptive economic management modes to design the model verification experiment. The concepts of EMA and BDT have been described in the previous sections. Accordingly, a new economic management model is implemented to optimize the business process by integrating business operations and fund management [27], thus providing information and fund as an effective enterprise service. Figure 9 describes the specific structure.

Afterward, the rationality and practicability of the proposed economic management model will be verified by predicting the enterprise's income, fund, and HR structure. The experimental results are shown below.

The new economic management model needs to build three matrices.

(i) User scoring matrix. It collects the user's specific itemized scores. The corresponding matrix element is zero when users skip scoring a certain item.

(ii) User feature attention matrix. It represents the user's preference for features.

(iii) Item feature quality matrix. It represents the feature level of the item.

The rationale of the proposed integrated economic management model is to make edge nodes smarter, such as acting as agents for consumers and providers, to improve content retrieval and distribution. On the one hand, the new economic management model can help consumers use the edge router as a fast content repository to meet consumers' requests and as an intelligent agent to request content from upstream nodes. On the other hand, it can help providers use optimized intranetwork recovery/retransmission to detect packet loss and even accelerate content distribution. The present work aims to improve the performance of edge

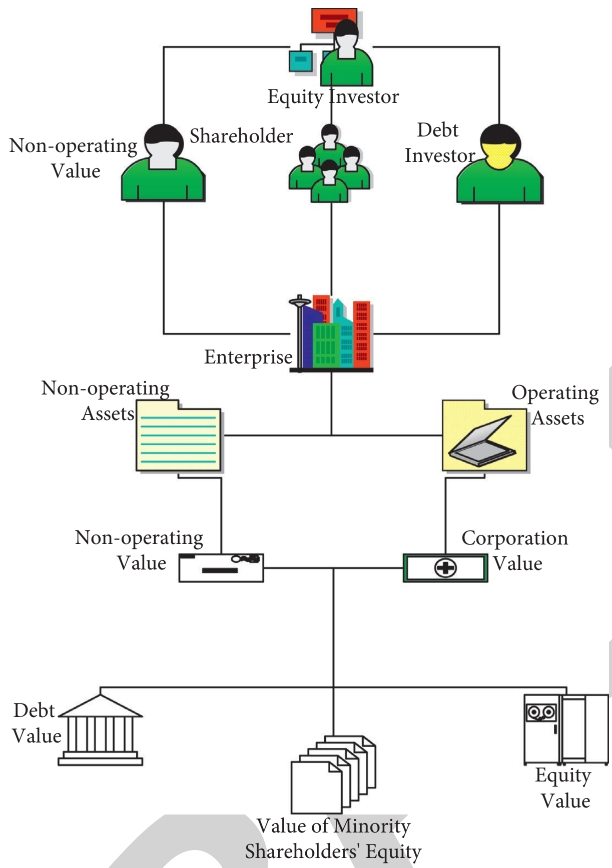


Figure 8: Relative valuation structure.

Table 4: Key features of the proposed integrated economic management model.

Key features	Meaning
Establishment of new ideas	Managers should abandon traditional ideas and learn innovative EEM knowledge to build a new management structure, especially the importance of the EEM. Inevitably, the optimized management model involves fund responsibility. Thus, it is necessary to strengthen trust between departments to ensure optimization
Centralized management of funds and budget	EEM should be supervised through formulation and approval. Enterprises must strictly adhere to the formulated articles to execute EEM

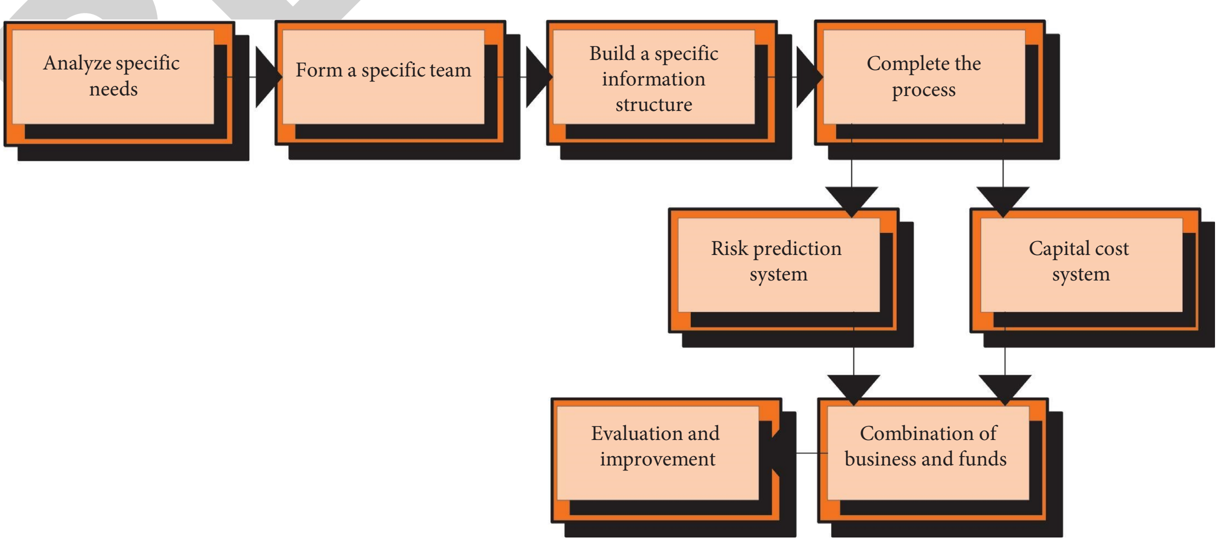


Figure 9: EEM structure.

networks. The simulation results show that the proposed economic management model can achieve efficient content retrieval and distribution and is more easily accessible to consumers and providers.

2.7. Enterprise Mobility Management (EMM). Mobile devices are blurring the line between work and personal life. Data show that 70% of the employees surveyed will use their smartphones or tablets to access the corporate network, and Small Office and Home Office (SOHO) has gradually become a new trend. Thus, mobile device management is challenging EMA [28].

As a fragment of enterprise mobility management (EMM), mobile device management might be the most widespread concern despite being a less critical factor. Indeed, a thorough understanding of EMM is the key to better managing mobile devices.

This section will take the EMM of ZIYA Information Company as an example to better explain the five key elements of EMM:

2.7.1. Mobile Device Management. Mobile device management manages the whole life cycle of mobile devices, such as system status, device information, call records, location analysis, and remote device inspection, to avoid mobile device failure.

2.7.2. Mobile Application Management. EMM-based unified application management can help access enterprise-sensitive data more securely. At the same time, mobile application management can manage the black-and-white list and forced installation of remote mobile application devices in combination with application policies.

2.7.3. Mobile Content Management (MCM). The enterprise shared files are protected through containerization or security sandbox technology. It can isolate, monitor, and control the distribution and access of sensitive information, prevent data from being transmitted, copied, and embezzled, and start an automatic protection mechanism for confidential files. In case of device loss or employee turnover, the administrator can remotely erase the device through the management background. Most enterprises prioritize mobile device security, especially in specific application scenarios, so MCM is essential for EMM.

2.7.4. Report Management. Report management realizes the statistics of application reports, equipment reports, policy reports, flow reports, and user reports. Meanwhile, it can display these statistic data in various forms, such as graphics, tables, and line charts, which can be exported in PDF, HTML, and Excel to facilitate management and control.

2.7.5. Security management. Security management secures enterprise mobile devices through data encryption, access authorization, shared devices, application packaging and control, device locking, and other measures. The specific functions are as follows:

Terminal security: if the device is lost or the Subscriber Identity Module (SIM) card is replaced by others, employees on the ZIYA self-service platform can easily set a password remotely to lock the device to prevent privacy disclosure. At the same time, they can remotely erase the data on the device and protect information security. Compliance management will notify the administrator immediately via e-mail or short message service (SMS) upon any user's operation violation.

Application security: It provides unified internal application release and management. Combined with the application strategy, the remote mobile device applications can be subjected to security management, such as application authentication, application function restriction, application time fence, black-and-white list, forced installation, and uninstallation of remote applications. Thereby, it ensures enterprise application security.

Transmission security: It establishes an encrypted transmission tunnel. It adopts a high-strength private one-machine-one-encryption algorithm to provide application-level address extension for enterprise applications, transmit personal, and enterprise applications separately, and ensure data transmission security.

Data security provides a safe sandbox container for enterprise applications to distinguish personal data and enterprise data in employee equipment. Also, it provides security protection against copy, screen capture, and data security watermark to ensure the local security of enterprise data in an all-around way.

Internet behavior security regulates employees' Internet behavior and restricts their accessible websites, protocols, and downloadable contents through a secure browser. Besides, it can configure browser home page and bookmarks and establish an enterprise's sensitive thesaurus. When sensitive words are input, users will be prompted to delete them. For smartphone-specific and SIM card-specific devices, Internet behavior security supports the binding of the cardholder. It can lock the mobile phone and SIM card after the employee changes the card privately.

A single function cannot meet the management needs of enterprises in the comprehensive management of enterprise mobile devices. With a more thorough understanding of the security and management of mobile devices, mobile applications, and data information, the devices can be better managed.

In the result part, a financial enterprise in a city is selected as the research object. The enterprise's annual OI is predicted using the proposed financial management model. The enterprise information, such as personnel, project type, and quantity, is used as the input as the prediction basis. Then, the predicted OI is compared with the enterprise's actual annual OI. Second, the enterprise's internal position and personnel deployment in the past year are estimated to verify the model's performance.

3. Actual Experimental Results and Analysis

3.1. Comparison between Annual OI Forecast and Actual Value. This section investigates some local enterprises' recruitment situations and position requirements. Then, the proposed EEM model predicts the enterprise annual OI based on enterprise funds, costs, and operations. The actual annual OI and the predicted value are compared to judge the feasibility and suitability of the proposed model. Table 5 calculates the specific results, and Figure 10 compares the outcomes. Subjective indexes can be both subjective and objective. Specifically, subjectivity refers to the individual judgment differences during the formation and measurement of subjective indexes, but it does not simply mean the measurement results come from subjective feelings. Surely, subjective indexes are designed, proposed, and measured by people with different features. Meanwhile, the subjective index conceptual design and measurement are inevitably affected by personal subjective factors. Such affection is reflected in index-mapping content and the index value deviations, which are consistent with the characteristics of the objective index.

Figure 10 suggests that the enterprise annual OI has continuously grown in the past six years, with a significant growth rate (GR) from 2018 to 2020. Noticeably, during 2015–2017, the predicted value was lower than the actual value. Presumably, at that time, the enterprise branches did not inform the headquarters of the sales situation timely. Therefore, in the headquarters-based prediction, the sales quota of the branches was not included in the calculation, resulting in a much lesser predicted value than the actual value. The largest error occurred in 2017, with a difference of 90,000 RMB. The smallest error occurred in 2015, with a deviation of 21,000 RMB. In 2018–2020, the predicted value was higher than the actual value; the maximum error occurred in 2019, with a difference of 2.762 million RMB; the minimum error occurred in 2018, with a deviation of 122,000 RMB. The model can predict the enterprise sales quota with high accuracy. However, in recent years, the sales quota of each enterprise has been growing steadily. If the sales quota drops suddenly, the model error might increase in the sales quota prediction, but the model can show the errors in time. Accordingly, enterprises can optimize and improve the sales strategy to avoid deficits.

On the other hand, Figure 10 implies that the predicted and actual growth trend is similar, indicating that the error is relatively small. Even if the maximum error reaches 2.762 million RMB, it is acceptably tiny compared with the overall 53.471 million RMB income. The error rate is only 276.2/5347.1 ≈ 5.1%. Therefore, the proposed DBT-based integrated EEM model can provide new ideas for EMA strategies.

Table 5: Comparison between predicted OI and actual OI (data unit: 10,000 RMB).

Year	Comparison items: Predicted value	Comparison items: Actual value
2015	306.8	308.9
2016	389.7	397.4
2017	901.8	910.8
2018	1859.7	1847.5
2019	3697.7	3421.5
2020	5449.8	5347.1

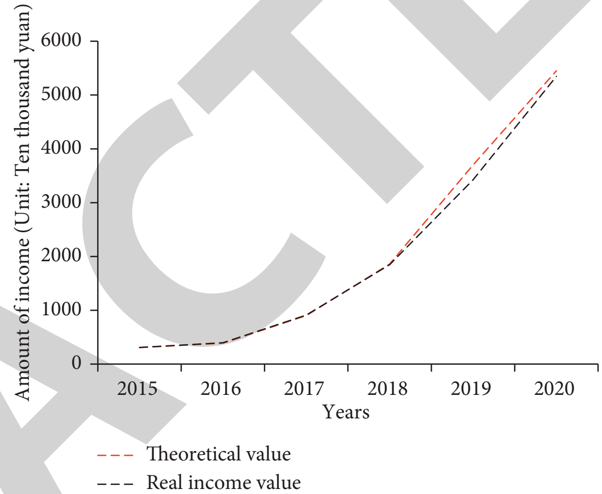


Figure 10: Comparison between predicted OI and actual OI.

3.2. Enterprise's HR Structure. When an enterprise reaches a certain scale, its internal positions will be divided into various types and levels. Whether a large-scale enterprise can carry out reasonable job planning for internal personnel will affect its smooth operation and management. Such factors include level, type, and the number of positions. In this section, the economic management model will be used to predict the type and number of different employees according to the enterprise type and size. The results will be compared with the actual number of employees to infer whether the inference of the model is reasonable and correct and further judge the rationality of the model. Figure 11 manifests the results.

The personnel allocation of four different enterprise organizations is forecasted, respectively, and the forecasted positions are operation strategy, sales development, and internal specification. It can be seen from the bar chart in Figure 11 that the predicted number of job deployments is very close to the actual number of employees in the enterprise.

In Figure 11(a), the minimum error is one person, and the maximum error is two. In Figure 11(b), the minimum error is two people, and the maximum error is six people. In Figure 11(c), the minimum error is one person, and the maximum error is eight persons. In Figure 11(d), the minimum error is one person, and the maximum error is two. Figure 11(c) is the primary organization with a large population base; thus, a two-people deviation in Figure 11(c) only has a tiny overall impact on the enterprise HR structure. By comparison, Figure 11(d) is a branch organization with a small population base, but the primary organization can allocate its personnel. Therefore, a two-people error still will not significantly impact its HR structure. Overall, the errors are within the acceptable range. Moreover, the predicted

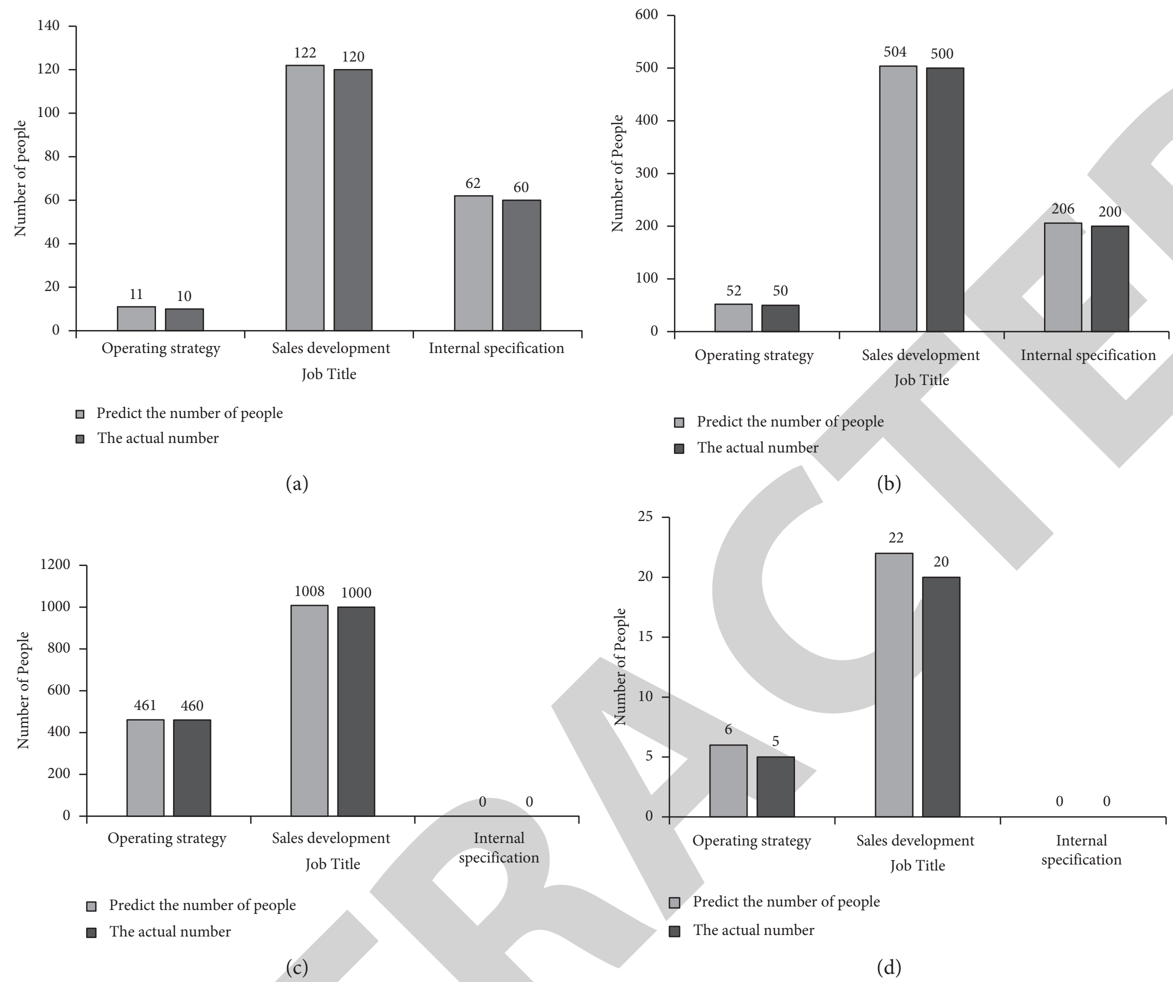


Figure 11: Comparison between the predicted number of positions and the actual number of positions ((a) data of level-2 enterprise organization, (b) data of level-1 enterprise organization, (c) data of level-1 enterprise branch office, and (d) data of level-2 enterprise branch office).

number of positions by the proposed integrated economic management model is higher than the actual number. Probably, it is because the proposed integrated economic management model is an open prediction method. Nevertheless, the performance of the proposed model is suitable, ensuring high accuracy for the deployment and calculation of the number of positions and personnel in the enterprise. The proposed model can apply well to innovative EEM. It can provide references for enterprise development by modifying the recruitment conditions timely and customizing recruitment strategies.

4. Conclusion

The recent years mark the emergence, evolution, and optimization of BDT and economic globalization, bringing new challenges and opportunities to financial enterprises. At the same time, the BDE has helped financial enterprises further digitalize economic management to adapt to the volatile economic environment. Therefore, the EEM model must be innovated through new technological means, such as BDT. As such, enterprises can gain a firm foothold in the fierce business competition and seek better development. To this end, this paper studies the innovation of EEM. Firstly, it studies and expounds on the BDT and EEM and integrates the two, and analyzes the influence of BDE on EEM. Following the hierarchical classification of the EEM, this paper explores the EEM innovation to provide the basis for pinpointing EMA. As a result, an integrated economic management model is proposed for a financial enterprise. Subsequently, the proposed model calculates the enterprise annual OI in the past six years based on historical data. The results imply that the proposed model shows a small error and can accurately predict the OI and cost data. Further, the proposed model estimates the position deployment in the four dimensions of the enterprise HR structure and compares the results with the actual number. Therefore, the proposed BDT-based integrated economic management model can broaden the EMA vision and provide theoretical bases for enterprise innovation. Lastly, some deficiencies are summarized, and the research prospect is envisioned. There are many objective laws in today's market economy. These laws aim to raise market economy proportion, maximize enterprise economic benefits, and optimize industrial clusters. These problems need to be gradually excavated. In view of this, the research and analysis on the objective economic law under the background of market economy visions a broad development prospect. The data analysis department holds detailed data and materials. Thus, they can conduct targeted research and analysis on the data content beneath the surface, excavate the hidden feature, and replace the perceptual development knowledge with rational knowledge. Furthermore, these analyses can sublime the understanding of objective economic laws and illustrate the enterprise development status and its internal relationship. On the one hand, enterprise managers and relevant departments can understand enterprise economic behaviors and enterprise development status and direction to improve EMA. On the other hand, full play should be given the data analysis in scientific enterprise operation, management, and decision-making. After analyzing the model errors causes, this paper believes that the following factors: the main development direction of the local city; professional types of local universities; and local natural factors are the most likely to deviate the model estimation. In the follow-up research, the model will be improved in these regards.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This study was supported by the Innovation and Development Research Project of Social Science in Anhui Province (2019CX073).

References

[1] Z. Ren, L. Shen, and Z. Zhou, “Research on the reform of digital economy professional training mode from the perspective of “Internet plus curriculum thinking”,” *Frontiers in Business, Economics and Management*, vol. 2, no. 3, pp. 32–35, 2021.

[2] Y. Liu, M. Peng, G. Shou, Y. Chen, and S. Chen, “Toward edge intelligence: multiaccess edge computing for 5G and Internet of Things,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6722–6747, 2020.

[3] J. L. Jimenez-Marquez, I. Gonzalez-Carrasco, J. L. Lopez-Cuadrado, and B. Ruiz-Mezcua, “Towards a big data framework for analyzing social media content,” *International Journal of Information Management*, vol. 44, pp. 1–12, 2019.

[4] V. Marinakis, H. Doukas, J. Tsapelas et al., “From big data to smart energy services: an application for intelligent energy management,” *Future Generation Computer Systems*, vol. 110, pp. 572–586, 2020.

[5] Y. Jia, X. Gao, and S. Julian, “Do firms use corporate social responsibility to insure against stock price risk? Evidence from a natural experiment,” *Strategic Management Journal*, vol. 41, no. 2, pp. 290–307, 2020.

[6] I. M. Matovic, “Combining agile and traditional project management as a better approach to project implementation,” in *Proceedings of the 20th International RAIS Conference*, vol. 1, NJ, USA, December 2020.

[7] A. Eizaguirre, M. García-Feijoo, and J. P. Laka, “Defining sustainability core competencies in business and management studies based on multinational stakeholders’ perceptions,” *Sustainability*, vol. 11, no. 8, p. 2303, 2019.

[8] O. Westlund, A. H. Krumsvik, and S. C. Lewis, “Competition, change, and coordination and collaboration: tracing news executives’ perceptions about participation in media innovation,” *Journalism Studies*, vol. 22, no. 1, pp. 1–21, 2021.

[9] M. Fajar, “Fair competition: the concept of regulation in the sharing economy,” *The Journal of Asian Finance, Economics and Business*, vol. 7, no. 11, pp. 637–645, 2020.

[10] S. Shao, Z. Hu, J. Cao, L. Yang, and D. Guan, “Environmental regulation and enterprise innovation: a review,” *Business Strategy and the Environment*, vol. 29, no. 3, pp. 1465–1478, 2020.

[11] R. Lin, Z. Xie, Y. Hao, and J. Wang, “Improving high-tech enterprise innovation in big data environment: a combinative view of internal and external governance,” *International Journal of Information Management*, vol. 50, pp. 575–585, 2020.

[12] J. Wang, Y. Yang, T. Wang, R. Sherratt, and J. Zhang, “Big data service architecture: a survey,” *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.

[13] M. Ghasemaghaei and G. Calic, “Does big data enhance firm innovation competency? The mediating role of data-driven insights,” *Journal of Business Research*, vol. 104, pp. 69–84, 2019.

[14] M. Ghasemaghaei and G. Calic, “Assessing the impact of big data on firm innovation performance: big data is not always better data,” *Journal of Business Research*, vol. 108, pp. 147–162, 2020.

[15] L. Tan, N. Shi, C. Yang, and K. Yu, “A blockchain-based access control framework for cyber-physical-social system big data,” *IEEE Access*, vol. 8, Article ID 77215, 2020.

[16] J. P. S. I. Hernández, B. Yañez-Araque, and J. Moreno-García, “Moderating effect of firm size on the influence of corporate social responsibility in the economic performance of micro-, small-and medium-sized enterprises,” *Technological Forecasting and Social Change*, vol. 151, Article ID 119774, 2020.

[17] Y.-K. Lin, C.-H. Huang, and S. G. Chen, “Applying network reliability in business management activities,” *IEEE Access*, vol. 9, Article ID 61532, 2021.

[18] S. Drobyazko, A. Barwińska-Małajowicz, B. Ślusarczyk, L. Zavidna, and M. D. Kropyvnytska, “Innovative entrepreneurship models in the management system of enterprise competitiveness,” *Journal of Entrepreneurship Education*, vol. 22, no. 4, pp. 1–6, 2019.

[19] M. I. Baig, L. Shuib, and E. Yadegaridehkordi, “Big data adoption: state of the art and research challenges,” *Information Processing & Management*, vol. 56, no. 6, Article ID 102095, 2019.

[20] A. Johnston, B. Segrestin, and A. Hatchuel, “From balanced enterprise to hostile takeover: how the law forgot about management,” *Legal Studies*, vol. 39, no. 1, pp. 75–97, 2019.

[21] M. Giannakis, R. Dubey, I. Vlachos, and Y. Ju, “Supplier sustainability performance evaluation using the analytic network process,” *Journal of Cleaner Production*, vol. 247, Article ID 119439, 2020.

[22] J. Yan, L. Feng, A. N. Steblyanskaya, and S. Fu, “Comparative study of discounted cash flow and energy return on

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9832516, 1 page
https://doi.org/10.1155/2022/9832516

Retraction

Retracted: Feature Extraction and Identification of Calligraphy Style Based on Dual Channel Convolution Network

Security and Communication Networks

Received 10 November 2022; Accepted 10 November 2022; Published 20 November 2022

Security and Communication Networks has retracted the article titled "Feature Extraction and Identification of Calligraphy Style Based on Dual Channel Convolution Network" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] Z. Wang, "Feature Extraction and Identification of Calligraphy Style Based on Dual Channel Convolution Network," *Security and Communication Networks*, vol. 2022, Article ID 4187797, 11 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4187797, 11 pages
https://doi.org/10.1155/2022/4187797

Research Article

Feature Extraction and Identification of Calligraphy Style Based on Dual Channel Convolution Network

Ziyan Wang

Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

Correspondence should be addressed to Ziyan Wang; wzyabbamm525@stu.xjtu.edu.cn

Received 25 February 2022; Revised 17 March 2022; Accepted 28 March 2022; Published 16 May 2022

Academic Editor: Chin-Ling Chen

To improve the effect of calligraphy style feature extraction and identification, this study proposes a calligraphy style feature extraction and identification technology based on two-channel convolutional neural network and constructs an intelligent calligraphy style feature extraction and identification system. Moreover, this paper improves the C3D network model and retains 2 fully connected layers. In addition, by extracting the outline skeleton and stroke features of calligraphy characters, this paper calculates the feature weight and authenticity determination function and constructs an authenticity identification system. The experimental study shows that the calligraphy style feature extraction and identification system based on the dual-channel convolutional neural network proposed in this paper has a good performance in calligraphy style feature extraction and identification.

1. Introduction

The digitized historical calligraphy works, books, and signatures contain tens of thousands of individual calligraphy characters, covering the original calligraphy works or the photocopies of various historical periods and famous calligraphers. These calligraphic works are displayedin the CADAL portal. Users can search these calligraphic works based on metadata information such as the title of the work and the name of the calligrapher. At the same time, some calligraphy characters can also be retrieved by content-based methods by inputting or drawing a calligraphy character image. However, for digital calligraphy works with a large amount of data, the original outline-based calligraphy character retrieval technology has gradually been unable to meet the needs of user applications because it takes too long. At the same time, in the actual service, the calligraphy character database needs to be continuously supplemented. When adding new calligraphy characters to the calligraphy character database, in order to obtain complete metadata, each calligraphy character needs to be manually identified and entered into the database by the administrator. This is a very time-consuming and labor-intensive job. In addition, with the continuous increase of digitized calligraphy works, users also put forward new demands for better use of these calligraphy characters. There are many users who wish to use the words of a famous calligrapher in these calligraphy resources to inscribe their study or company. However, among those characters specified by the user, there may be some characters that are not present in the existing works of the individual specified calligraphers. There is a need for a technique that can generate such calligraphic characters.

When identifying calligraphy samples through CNN, we must first extract the features of the image. The particularity of calligraphy samples is reflected in two aspects: the writing background is simple, usually black and white. The various writing styles are not completely unique, but there is a mutual reference. Moreover, there are many similar properties in the features. Before the calligraphy font is recognized, a good feature extractor is needed to extract the features of the samples [1]. In the Caffe neural network framework, the LMDB data storage format is used to store images, so it is necessary to format the training samples and test samples. The convolutional neural network is mainly composed of three parts: a feature extraction layer, a fully connected layer, and a classifier [2]. Among them, the feature extraction layer performs feature extraction on the input samples, and the obtained feature images are subjected to multiple visual enlargement processing to retain the representative image feature information, and then the features are vectorized through the fully connected layer. Finally, the retained eigenvectors are probability normalized by the classifier, and the categories are divided according to the principle of maximum probability [3].

This paper proposes a calligraphy style feature extraction and identification technology based on two-channel convolutional neural network, constructs an intelligent calligraphy style feature extraction and identification system, and improves the effect of calligraphy style feature extraction and identification.

2. Related Work

Reference [4] proposed a calligraphic character retrieval method based on the similarity of calligraphic character outlines. This method was used to test 20 calligraphic characters, and the precision was close to 90% under the 90% recall rate. Reference [5] proposed a method for retrieving calligraphy characters based on the skeleton features of calligraphy characters. Literature [6] proposes a fast multilayer retrieval method for calligraphy characters, which improves the retrieval speed through two-layer calligraphy character retrieval. The full-rate and precision rate remain basically unchanged. Reference [7] proposes a SC.HoG feature to describe calligraphy characters, which expresses the position information of a certain contour point of calligraphy characters and the distribution information of contour points around the contour point, so as to perform shape-based retrieval for calligraphy characters. Reference [8] first performs pruning processing, and filters out calligraphic character images that are impossible to be similar to the calligraphic character image to be retrieved by comparing the complexity of calligraphy characters, stroke density, and other characteristics. Dimensional calligraphy image feature data is used for retrieval, and PK-tree is used to improve retrieval speed. Reference [9] proposed an ancient book content retrieval method based on visual similarity. The above-mentioned content-based calligraphy character retrieval research shows that the extracted calligraphy characters have high feature dimensions. Therefore, with the continuous increase of calligraphy characters in the database, in the retrieval of large data calligraphy characters, when comparing and calculating the similarity, it still consumes a lot of time, and the retrieval speed becomes slower and slower, which cannot meet the needs of online services.

Reference [10] proposes a high-dimensional calligraphic character index method based on hybrid distance-tree (HD-Tree) to speed up retrieval. Given a query calligraphic character image, the hybrid distance tree index is used to complete high-dimensional calligraphy characters, image query. Reference [11] proposed an interactive high-dimensional vector indexing method based on partial-distance-map (PDM). PDM mainly establishes the relationship between the concept of Chinese character semantic level and the underlying shape features through user feedback to improve the retrieval speed of the massive calligraphy character database.

Literature [12] proposed a calligraphy character recognition method, which is based on the calligraphy character retrieval. First, the calligraphy character skeleton similarity retrieval method is used to retrieve the calligraphy in the database that is similar to the calligraphy character image to be recognized. Then, according to the semantic annotation of the retrieved similar calligraphic character images in the database, the recognition results are given. In the experiment, 300 calligraphy characters were tested and the recognition rate reached 96.3%. This method is based on retrieval, and the recognition time depends on the time required to retrieve similar calligraphic characters in the database. When the database is large, the time efficiency of this recognition method is very low. The polygon approximation method is to use polygon line segments to approximate the shape edge, and the minimum error is generally used to measure the quality of the approximation method used. Reference [13] proposed a shape polygon approximation method based on Hopfield neural network. Reference [14] proposed a relaxed iterative matching algorithm. This method uses the start and the end coordinates, length, and direction to describe stroke outline segments. A spline refers to a piecewise-defined polynomial parametric curve. For a given node vector, all splines of degree n form a vector space. A basis for this space is a B-spline of degree n. Reference [15] proposed a B-spline curve matching method with affine invariance. The commonly used scale spaces are: Gaussian scale space, wavelet scale space, and shape scale space. The salient feature points of the target refer to those feature points that still exist in the simplified representation by tracking the positions of the feature points at different scales to give a simplified form of the shape.

The detection and description of image local features can be effectively used to identify objects. SIFT features are based on the principle that the key points of the local appearance on the object are independent of the size and rotation of the image. In addition, the tolerance of SIFT features to changes in light, noise, and viewing angle is also higher [16]. Based on the above advantages, the feature information of the image is relatively easy to extract. After the image feature database is established, it is easy to identify objects in the feature database, and the recognition rate is high. SIFT feature description also has a high detection rate for objects with partial occlusion, and only a small amount of SIFT image features can be used to calculate the position and orientation of key points. The SIFT method has high operating efficiency, and the SIFT feature library has a large amount of information, and the recognition speed is close to real-time operation, so the algorithm is suitable for efficient and accurate matching of massive data [17].

The speeded-up robust features (SURF) operator is an improved algorithm proposed on the basis of maintaining the excellent performance of the SIFT operator. SURF solves the shortcomings of high computational complexity and time-consuming SIFT [18]. The extraction and its feature vector description have been improved, and the calculation speed has been greatly improved. SURF uses the Hessian matrix to stably obtain the extreme value of the image space. However, in the stage of finding the main direction, it completely depends on the gradient direction of the local pixel points, which may cause the main direction to be inaccurate [19], and because the feature vector extraction and matching are completely dependent on the above, the main direction is determined by the step and even a small deviation angle may cause the error of feature matching [20].

3. Two-Channel Convolutional Network Image Feature Extraction

The training idea of the convolutional neural network is to define a loss function, train the network through the existing samples, optimize the network parameters by using the back-propagation algorithm, and obtain the parameters of the network when the loss function is the smallest.

The neural network defines the loss function for a single sample (x, y) as

$$J(W, b; x, y) = \frac{1}{2}\left\| h_{w,b}(x) - y \right\|^2. \tag{1}$$

Here, W and b are the weight and bias parameters of the network respectively, $h_{w,b}(x)$ is the output of the sample x through the neural network, and y is the expected output value of the sample.

For a training set $\{(x^{(1)}, y^{(1)}), \cdots, (x^{(m)}, y^{(m)})\}$ with m samples, the overall loss function is

$$\begin{aligned} J(W,b) &= \left[\frac{1}{m}\sum_{i=1}^{m} J\left(W,b;x^{(i)},y^{(i)}\right)\right] + \frac{\lambda}{2}\|W\|^2, \\ &= \left[\frac{1}{M}\sum_{i=1}^{m}\frac{1}{2}\left\|h_{w,b}\left(x^{(i)}\right)-y^{(i)}\right\|^2\right] + \frac{\lambda}{2}\sum_{l=1}^{nl-1}\sum_{i=1}^{Sl}\sum_{j=1}^{Sl+1}\left(W_{ji}^{l}\right)^2. \end{aligned} \tag{2}$$

Here, the first term is the mean square error, the second term is the weight decay term, which is used to prevent overfitting, n_l is the number of network layers, and s_l is the number of neurons contained in the l-th layer network.

Before training the network, the algorithm first randomly initializes the parameters of the weight W and bias b of the network and uses the gradient descent method to update W and b according to the following equations during training:

$$W^{(l)} = W^{(l)} - \alpha \frac{\partial}{\partial W^{(l)}} J(W, b), \tag{3}$$

$$b^{(l)} = b^{(l)} - \alpha \frac{\partial}{\partial b^{(l)}} J(W, b). \tag{4}$$

Among them, α is the learning rate, which determines whether the loss function can converge to the local minimum and the convergence speed. To update the parameters, it is necessary to obtain the partial derivatives of $J(W, b)$ with respect to W and b, which uses the backpropagation algorithm, whose calculation formulas are shown in

$$\frac{\partial}{\partial W^{(l)}} J(W, b) = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial W^{(l)}} J\left(W, b; x^{(i)}, y^{(i)}\right)\right] + \lambda W^{(l)}, \tag{5}$$

$$\frac{\partial}{\partial b^{(l)}} J(W, b) = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial b^{(l)}} J\left(W, b; x^{(i)}, y^{(i)}\right)\right]. \tag{6}$$

According to the chain method, it is derived as

$$\begin{aligned} &= \left(\frac{\partial J(W, b; x, y)}{\partial z^{(l+1)}}\right)^T \cdot \frac{\partial z^{(l+1)}}{\partial W^{(l)}}, \\ \frac{\partial J(W, b; x, y)}{\partial W^{(l)}} &= \left(\frac{\partial J(W, b; x, y)}{\partial z^{(l+1)}}\right)^T \cdot \frac{\partial\left[W^{(l)} a^{(l)} + b^{(l)}\right]}{\partial W^{(l)}}, \\ &= \left(\frac{\partial J(W, b; x, y)}{\partial z^{(l+1)}}\right)^T \cdot a^{(l)}. \end{aligned} \tag{7}$$

If the residual of the l-th layer network is $\delta^{(l)} = \partial J(W, b; x, y)/\partial z^{(l)}$, then the residual of the output layer (n_l-th layer) is

$$\begin{aligned} &= \frac{\partial J(W,b;x,y)}{\partial z^{(n_l)}} = \frac{\partial}{\partial z^{(n_l)}}\frac{1}{2}\left\|h_{W,b}(x) - y\right\|^2, \\ \delta^{(n_l)} &= \frac{\partial}{\partial z^{(n_l)}}\frac{1}{2}\left\|f\left(z^{(n_l)}\right) - y\right\|^2 = \left(f\left(z^{(n_l)}\right) - y\right) \cdot f'\left(z^{(n_l)}\right), \\ &= \left(a^{(n_l)} - y\right) \cdot f'\left(z^{(n_l)}\right). \end{aligned} \tag{8}$$

Using mathematical induction, the residual of the l-th layer is obtained, and the expression is

$$\delta^{(l)} = \frac{\partial J(W, b; x, y)}{\partial z^{(l)}} = \left(w^{(l)}\right)^T \delta^{(l+1)} \cdot f'\left(z^{(l)}\right). \tag{9}$$

Therefore, the partial derivative of $J(W, b)$ with respect to W is obtained as

$$\frac{\partial J(W, b; x, y)}{\partial W^{(l)}} = \left(\delta^{(l+1)}\right)^T \cdot a^{(l)}. \tag{10}$$

Similarly, the partial derivative of $J(W, b)$ with respect to b can be obtained as

$$\frac{\partial J(W, b; x, y)}{\partial b^{(l)}} = \delta^{(l+1)}. \tag{11}$$

After the partial derivatives of $J(W, b)$ with respect to W and b are obtained, the parameters of the weight W and the bias b can be updated and optimized according to (3) and (4).

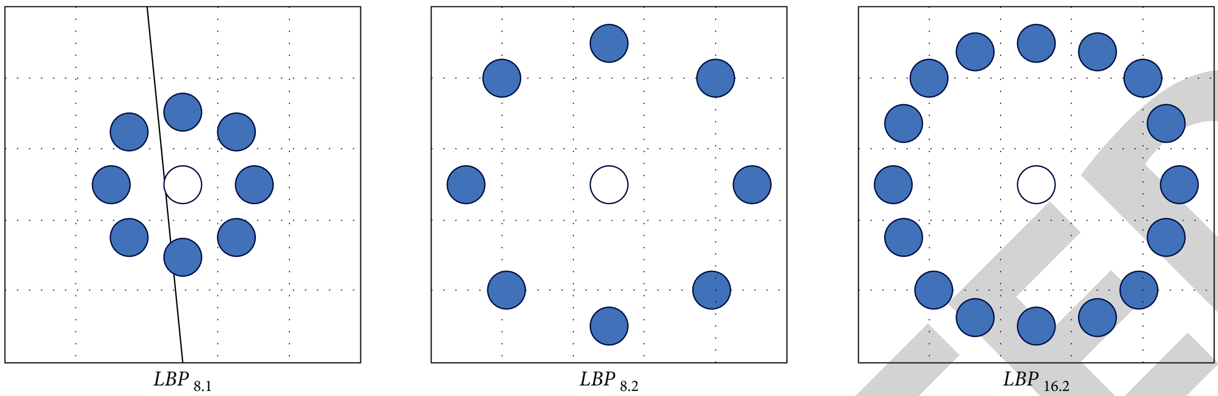


Figure 1: Several common LBP operators.

Due to the introduction of convolutional layers and pooling layers, the training of CNN cannot be directly applied to the training method of DNN. There are several differences between them:

(1) There is no activation function in the pooling layer, which can be solved by setting the activation function of this layer to $f(z) = z$

(2) The pooling layer compresses the feature map during forward propagation

(3) DNN is a fully connected network, the output of the current layer is directly obtained by matrix multiplication, and the convolution layer is obtained by summing several matrices to obtain the current output

(4) For the convolutional layer, the operation method used by W is convolution

In view of these differences, the backpropagation algorithm of CNN is discussed in the following two situations:

(1) Knowing the $\delta^{(l)}$ of the pooling layer, the $\delta^{(l-1)}$ of the previous hidden layer is derived.

The pooling layer first restores $\delta^{(l)}$ all submatrices to the size before pooling during backpropagation. If maximum pooling is used, the value of the $\delta^{(l)}$ submatrix corresponding to the pooling area is placed in the position where the maximum value is taken during the forward propagation. If average pooling is used, the values of the pooled area corresponding to the $\delta^{(l)}$ submatrix are averaged and put into the restored matrix. This restoring process is defined as *upsample*. The restoration process is further explained below with an example.

We assume that the pooling window size is 22, the window sliding step is 2, and the k-th submatrix of $\delta(l)$ is $\delta_k^{(l)} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$, and the submatrix is restored to the size before pooling and becomes $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 6 & 8 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. If it is maximum pooling, we assume that the maximum positions are the lower left, lower right, upper right, and lower right during forward propagation, then the restored matrix is $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 4 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$. If it is average pooling, the matrix after reduction is $\begin{pmatrix} 0.5 & 0.5 & 1 & 1 \\ 0.5 & 0.5 & 1 & 1 \\ 1.5 & 1.5 & 2 & 2 \\ 1.5 & 1.5 & 2 & 2 \end{pmatrix}$.

the value of the upper layer $\partial J(W, b; x, y)/\partial a_k^{(l-1)}$ can be obtained by upsampling, then there are

$$\delta_k^{(l-1)} = \frac{\partial J(W, b; x, y)}{\partial z_k^{(l-1)}} = \frac{\partial J(W, b; x, y)}{\partial a_k^{(l-1)}} \frac{\partial a_k^{(l-1)}}{\partial z_k^{l-1}}, \quad (12)$$

$$\text{upsample}\left(\delta_k^{(l)}\right) \cdot f'\left(z^{(l-1)}\right).$$

For the tensor $\delta^{(l-1)}$, we have

$$\delta^{(l-1)} = \text{upsample}\left(\delta^{l}\right) \cdot f'\left(z^{(l-1)}\right). \quad (13)$$

(2) Knowing the $\delta^{(l)}$ of the convolutional layer, it is deduced to $\delta^{(l-1)}$ of the previous hidden layer.

In DNN, the recurrence formula of $\delta^{(l)}$ and $\delta^{(l-1)}$ is

$$\delta^{(l-1)} = \frac{\partial J(W,b;x,y)}{\partial z^{(l-1)}} = \frac{\partial J(W,b;x,y)}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial z^{(l)}} = \delta^{(l)} \cdot \frac{\partial z^{(l)}}{\partial z^{(l-1)}}. \quad (14)$$

According to the forward propagation formula of the convolutional layer, we get

$$z^{(l)} = a^{(l-1)} * W^{(l)} + b^{(l)} = f\left(z^{(l-1)}\right) * W^{(l)} + b^{(l)}. \quad (15)$$

Then, there are

$$\delta^{(l-1)} = \delta^{(l)} \cdot \frac{\partial z^{(l)}}{\partial z^{(l-1)}} = \delta^{(l)} * \text{rot180}\left(W^{(l)}\right) \cdot f'\left(z^{(l-1)}\right). \quad (16)$$

Among them, $rot180()$ means to rotate 180 degrees.

Now, given the $\delta^{(l)}$ of the convolutional layer, the gradient of this layer W is

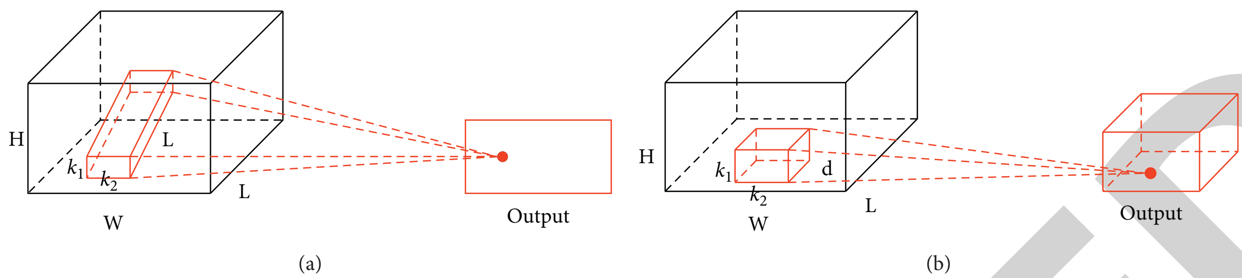


FIGURE 2: Convolution to generate feature map: (a) feature map generated by 2D convolution of the video frame sequence; (b) the video frame sequence is subjected to 3D convolution to generate a feature map.

$$\frac{\partial J(W,b;x,y)}{\partial W^{(l)}} = \frac{\partial J(W,b;x,y)}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial W^{(l)}} = a^{(l-1)} * \delta^{(l)}. \tag{17}$$

Since $\delta^{(l)}$ is a three-dimensional tensor and b is a one-dimensional vector, the algorithm sums the submatrix items of $\delta^{(l)}$ respectively, and the obtained one-dimensional error vector is the gradient of b. The calculation formula is

$$\frac{\partial J(W,b;x,y)}{\partial b^{(l)}} \sum_{u,y} \left(\delta^{(l)}\right) u, y. \tag{18}$$

The original LBP operator is defined in a fixed rectangular neighborhood of 3×3, which cannot meet the needs of textures of different sizes, so a circular LBP operator is generated. Compared with the original LBP operator, the circular LBP operator has two improvements: (1) the fixed neighborhood of 33 is extended to an arbitrary neighborhood; (2) the square neighborhood is no longer used, and the circular neighborhood is adopted. The circular LBP operator can be expressed as *LBPP*, *R*, where *R* is the radius of the circular neighborhood and *P* is the number of sampling points. Several common LBP operators are shown in Figure 1.

The calculation formula of the circular LBP operator is

$$\mathrm{LBP}_{P,R} = \sum_{i=1}^{P} s(g_i - g_c) 2^{i-1}. \tag{19}$$

Here, $s(u) = \begin{cases} 1, u \geq 0 \\ 0, \text{other} \end{cases}$, g_c is the gray value of the pixel in the center of the neighborhood, and g_i is the gray value of the pixel of the i-th sampling point in the neighborhood. To calculate the pixel gray value of the i-th sampling point, the coordinates (x_i, y_i) of the i-th sampling point are first calculated, and the calculation formula is

$$\begin{cases} x_i = x_c + R \cos\left(\dfrac{2\pi i}{P}\right), \\ y_i = y_c + R \sin\left(\dfrac{2\pi i}{P}\right). \end{cases} \tag{20}$$

Here, *(xc, yc)* is the coordinate of the center pixel. It can be seen from Figure 1 that some sampling points on the boundary of the circular neighborhood cannot just fall within the pixel grid, and may also fall on the boundary. At this time, the pixel value of the sampling point is calculated by the bilinear interpolation method, and the calculation formula is

$$g(x_i, y_i) = [1 - x_i x_i] \begin{bmatrix} g(x_{i-1}, y_{i-1}) & g(x_{i-1}, y_i) \\ g(x_i, y_{i-1}) & g(x_i, y_i) \end{bmatrix} \begin{bmatrix} 1 - y_i \\ y_i \end{bmatrix}. \tag{21}$$

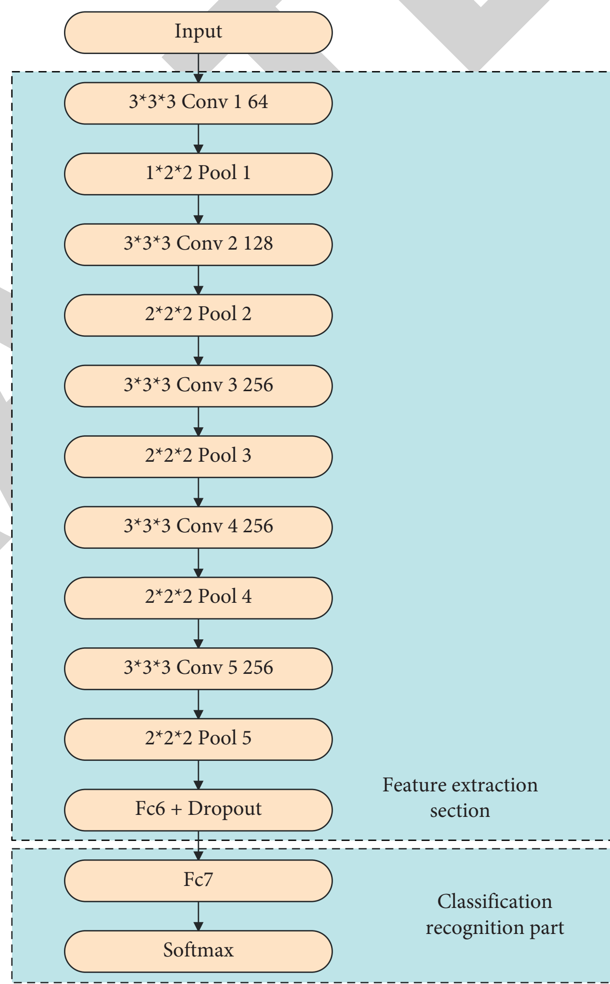


FIGURE 3: C3D network structure diagram.

In order to reduce the variety of binary patterns, Ojala proposes an equivalent pattern, which is also called uniform LBP. The main idea is that if the binary string corresponding to the mode does not jump from 0 to 1 or from 1 to 0 more

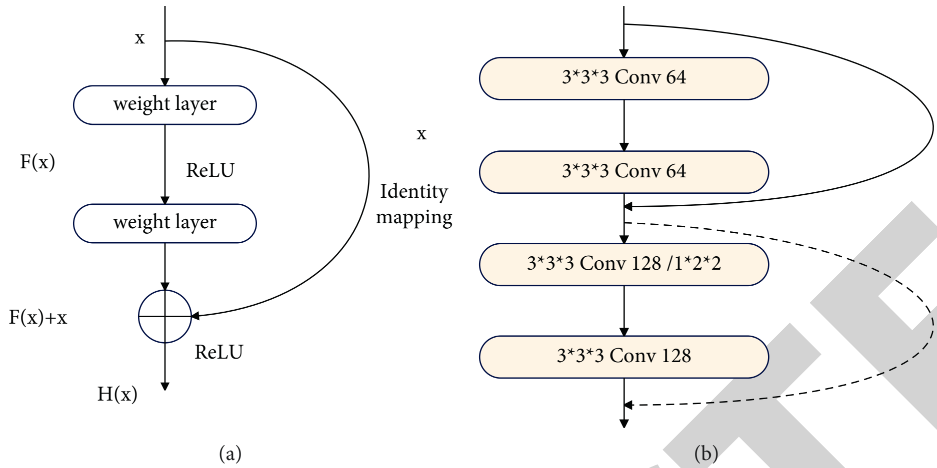


Figure 4: Residual network: (a) residual structure; (b) residual module structure.

than twice, then the mode is an equivalent mode, and all other modes are mixed modes, which are expressed as

$$U\left(\mathrm{LBP}_{p,R}\right)=\sum_{i=1}^{p}\left|s\left(g_i-g_c\right)-s\left(g_{i-1}-g_c\right)\right|. \tag{22}$$

Here, $g_0=g_p$, and the mode satisfying $U\leq 2$ is the equivalent mode. The improved mode types are reduced from the original 2^{P} to $P(P-1)+2$, which not only does not lose image information, but also reduces the impact of high-frequency noise.

Three-dimensional convolution (3D convolution) performs convolution operations on both the spatial and temporal domains by extending the convolution kernel to the temporal domain. In this way, it can extract both spatial information of a single frame of image and temporal information between adjacent video frames. 3D convolution combines multiple consecutive video frame images into a cube and slides the 3D convolution kernel on the cube to perform the convolution operation. When performing the convolution operations on a sequence of video frames, the difference between 2D convolution and 3D convolution is as follows: after a 2D convolution, the generated feature map is an image, which loses the time information of the input data. However, through 3D convolution, the generated feature map is still a sequence of feature maps, which effectively captures the motion information of the target. The convolution process of 2D convolution and 3D convolution on the video frame sequence is shown in Figure 2.

The network used in this section improves the original C3D network model and retains two fully connected layers, and its network structure is shown in Figure 3. The improved C3D network contains 5 convolutional layers, 5 pooling layers, 2 fully connected layers, and 1 Softmax classifier. A ReLU layer is added after each convolutional layer and the first fully connected layer in the network. This is because there is a simple linear relationship between the layers of the network. By introducing a nonlinear function as an activation function, the algorithm increases the nonlinear relationship between the layers so that the network can fit

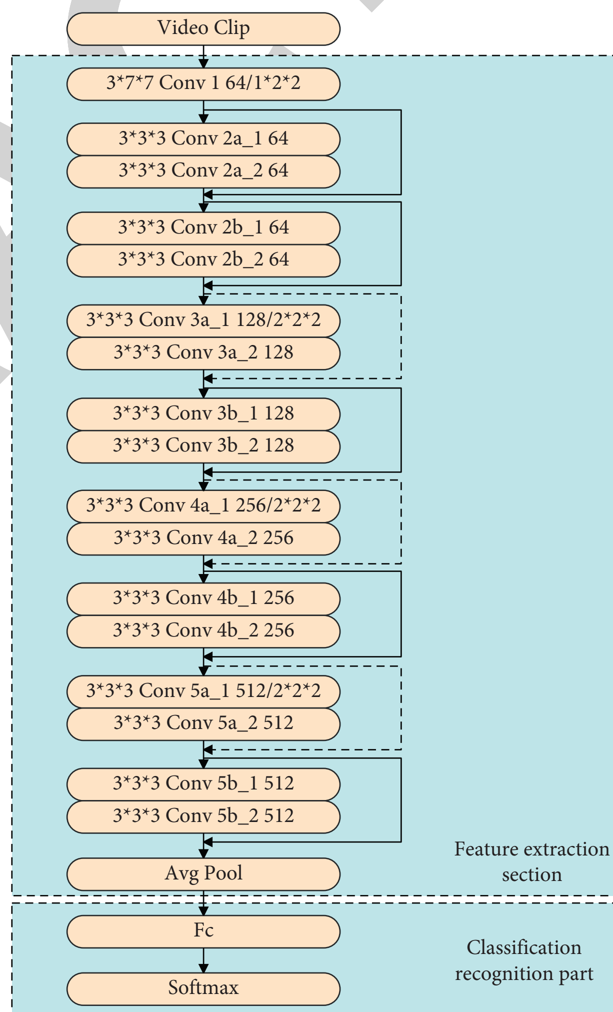


Figure 5: R3D network structure diagram.

complex functions. The formula of the ReLU function is shown in equation (5).

$$f(x)=\max(0,x). \tag{23}$$

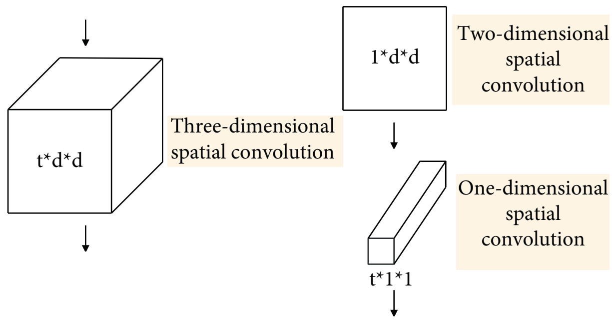


Figure 6: 3D convolution decomposition process.

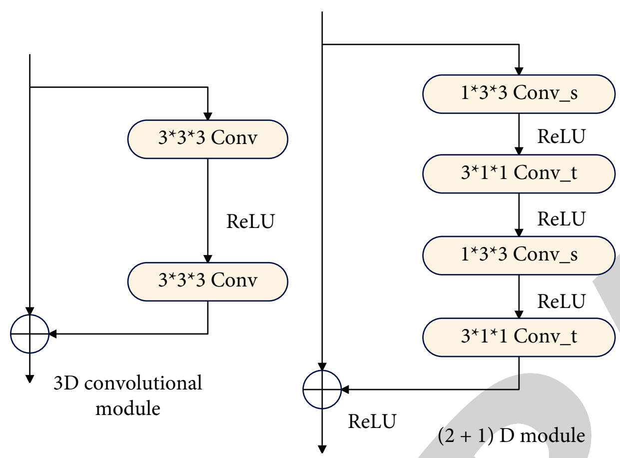


Figure 7: $(2+1)$ D decomposition process.

ResNet adopts a residual structure, as shown in Figure 4(a). This module has two branches, one of which is the normal convolutional layer output and the other that directly connects the input to the output. The final output of the module is the arithmetic sum of the two branches, which is formulated as

$$H(x) = F(x) + x. \tag{24}$$

Here, $H(x)$ represents the output of the entire structure, x is the input, and $F(x)$ is the output of the convolutional layer. ResNet defines a residual function $F(x) = H(x) - x$. When all parameters in the $F(x)$ branch are 0, $H(x) = x$ is the identity map. ResNet no longer learns the output of the entire structure, it is changed to learn the difference between the target value $H(x)$ and the input x, and the training goal is to make the residual function $F(x)$ approach 0.

Obviously, fitting the residual function is easier than fitting the identity mapping function, and the training process of the network is optimized. The residual structure is implemented by connecting the forward neural network and the identity map without introducing additional parameters. This will not increase the computational complexity of the network, and the network training method still uses backpropagation. After adopting the residual structure, the deep neural network achieves the ideal classification effect. A typical residual module structure is shown in Figure 4(b), which consists of stacking

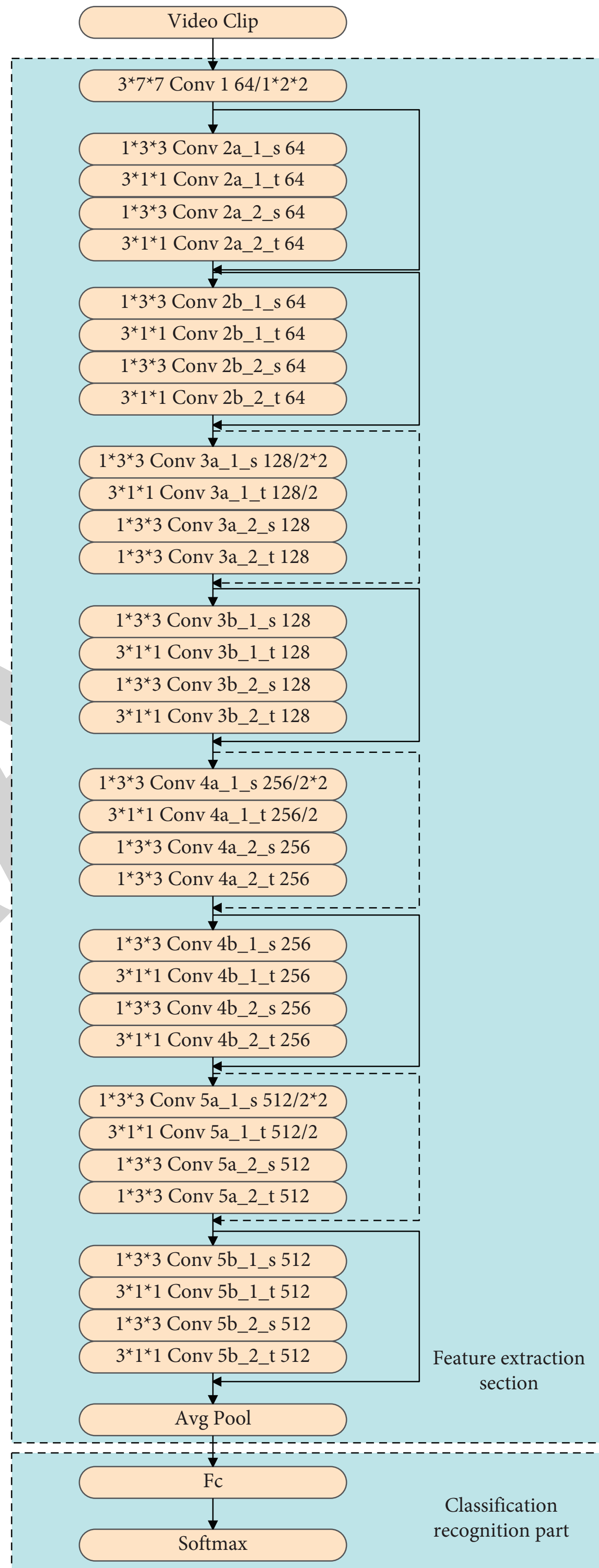


Figure 8: Structure diagram of R $(2+1)$ D network.

convolutional layers. Among them, $3 \times 3 \times 3$ is the size of the convolution kernel of this layer, 128 is the number of convolution kernels, and $/1 \times 2 \times 2$ indicates that the

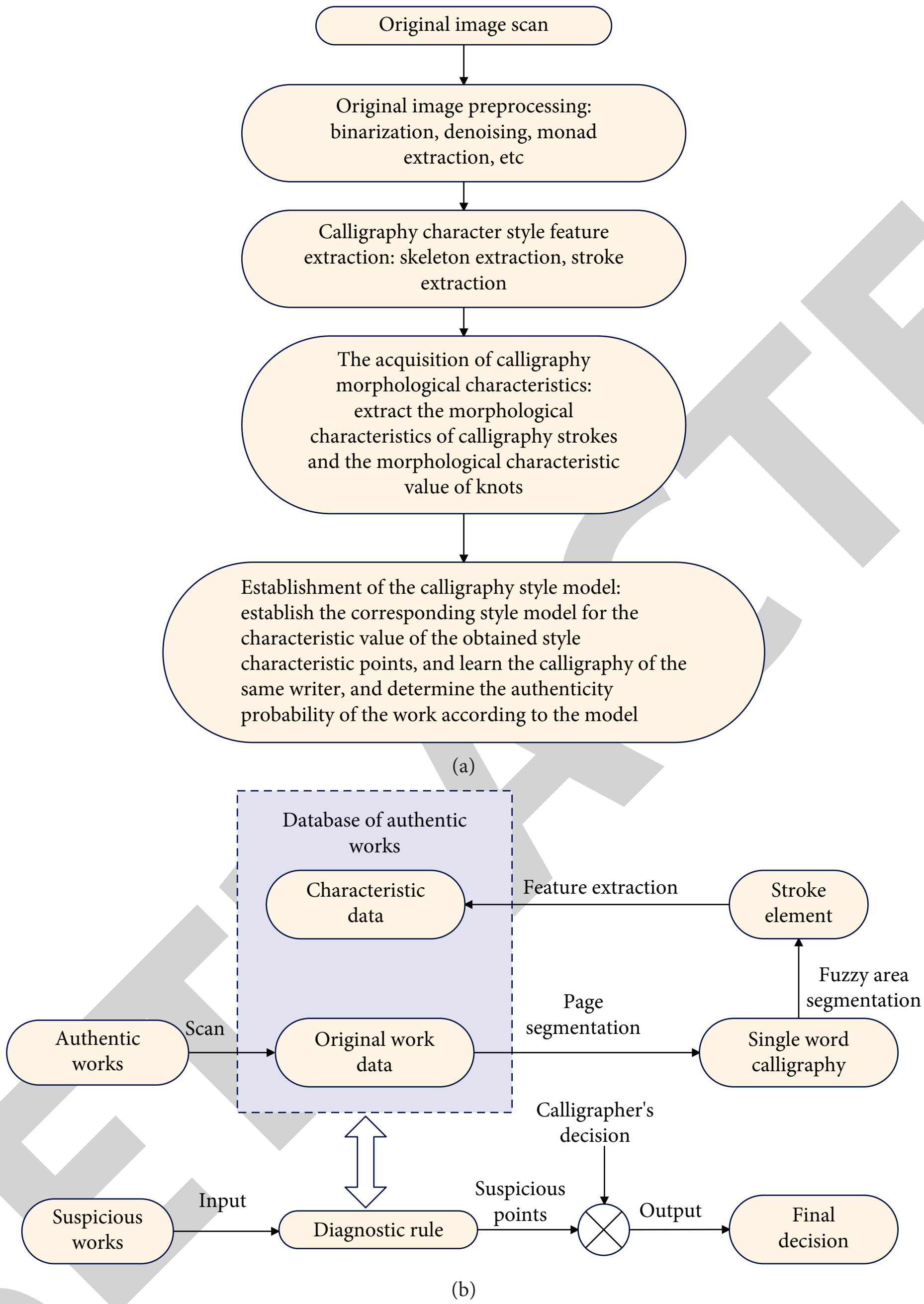


FIGURE 9: System design diagram. (a) Calligraphy identification block diagram. (b) System structure diagram.

convolution kernel moves with a step size of 1 in the time domain and 2 in the spatial domain.

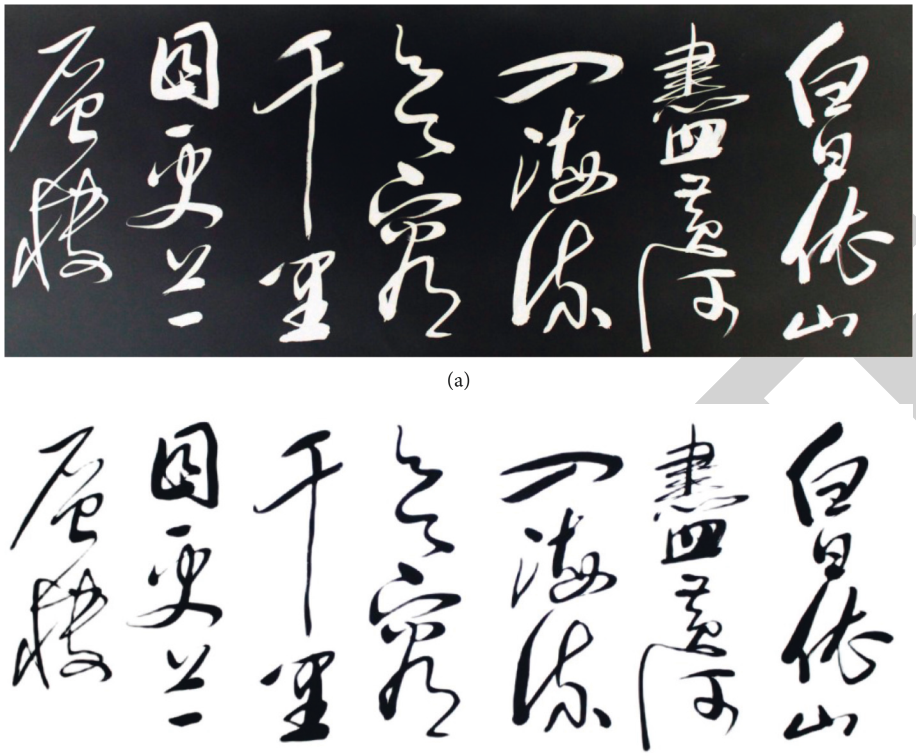
It is usually implemented by 1 × 1 convolution, and the formula is

$$H(x) = F(x) + Wx. \tag{25}$$

It should be noted that scheme (1) is not applicable to the case where the output feature map of the convolutional layer shown in Figure 4 is halved with respect to the size of the identity map feature map and the number of channels is doubled. Therefore, the dashed connection of the identity mapping in the R3D network structure adopted in this paper adopts the method in scheme (2). By adding a convolutional layer with a convolution kernel size of $1 \times 1 \times 1$ and a convolution stride of $1 \times 2 \times 2$ or $2 \times 2 \times 2$, the input feature map size and number of channels of the two branches are kept consistent.

The structure of the R3D model used in this paper is shown in Figure 5.

The 3D convolution operation simultaneously extracts video spatial and temporal dimensions through a 3D convolution kernel. According to the three-dimensional convolution properties, spatial modeling and temporal modeling can be decomposed into two separate steps, which are replaced by two-dimensional spatial convolution (2D convolution) and one-dimensional temporal convolution (1D convolution).



(a)

(b)

Figure 10: Example of calligraphy image recognition. (a) Original calligraphy image. (b) Feature extraction image.

This process is called three-dimensional convolution decomposition, which is also called $(2+1)$D decomposition vividly.

The 3D convolutional network preserves both temporal and spatial information through layer-by-layer transfer. In this paper, a video clip is fed into a 3D convolutional neural network. We assume that the tensor generated by the i-th convolutional layer is z_i, then z_i is a four-dimensional tensor of size $N_i \times L \times H_i \times W_i$, where N_i is the number of convolution kernels of the i-th convolutional layer, L is the time dimension of the feature map, and $H_i \times W_i$ is the spatial dimension of the feature map. Each convolution kernel is a four-dimensional tensor of size $N_{i\text{-}1} \times t \times d \times d$, where N_{i-1} is the number of convolution kernels of the $i-1$-th convolutional layer, t is the size of the three-dimensional convolution time dimension, and $d \times d$ is the size of the three-dimensional convolution space dimension.

The $(2+1)$D decomposition is a decomposition of a three-dimensional convolution kernel with N_i tensor size $N_{i-1} \times t \times d \times d$ into a two-dimensional spatial convolution kernel with M_i tensor size $N_{i-1} \times 1 \times d \times d$ and a one-dimensional temporal convolution kernel with N_i tensor size $M_i \times t \times 1 \times 1$. The hyperparameter M_i determines the subspace dimension of the feature map between spatial convolution and temporal convolution. In order to keep the parameters of the network before and after decomposition unchanged, the calculation formula of Mi is as follows:

$$M_i = \left\lfloor \frac{td^2 N_{i-1} N_i}{d^2 N_{i-1} + tN_i} \right\rfloor. \tag{26}$$

Figure 6 shows the 3D decomposition process when the input tensor z_{i-1} is a single channel (that is, z_{i-1}). If the 3D convolution has a span in space or time (implementing downsampling), it should be decomposed in the spatial and temporal dimensions accordingly.

Compared with the 3D convolution, the $(2+1)$ D decomposed convolution module does not reduce the number of parameters, but a ReLU layer is added between the 2D convolution and the 1D convolution. This leads to an increase in the number of nonlinear functions in the network, allowing the network to fit more complex functions. The above-mentioned $(2+1)$ D decomposition process does not change the number of parameters of the network, and the server memory requirements are relatively high when training the network.

This paper improves on the above $(2+1)$ D decomposition. Specifically, this paper directly decomposes N_i three-dimensional convolution kernels of size $t \times d \times d$ into N_i two-dimensional spatial convolution kernels of size $1 \times d \times d$ and N_i one-dimensional temporal convolution kernels of size $t \times 1 \times 1$. The improved $(2+1)$ D decomposition can greatly reduce the network parameters and speed up the network operation.

The R $(2+1)$D network used in this paper is a network formed by decomposing the convolutional layer according

Table 1: Performance verification of the calligraphy style feature extraction and identification system based on the two-channel convolutional neural network.

Number	Feature extraction	Calligraphy identification	Number	Feature extraction	Calligraphy identification
1	95.40	88.44	23	92.23	92.90
2	93.31	91.18	24	92.61	90.85
3	92.42	91.00	25	91.36	86.01
4	93.74	90.42	26	95.78	89.15
5	91.26	91.14	27	95.51	92.90
6	91.88	92.32	28	92.04	92.23
7	95.84	91.39	29	90.57	87.78
8	94.49	87.94	30	90.95	92.01
9	90.50	88.17	31	90.37	89.01
10	95.33	92.67	32	91.67	88.79
11	90.11	87.35	33	90.72	91.06
12	93.07	87.78	34	92.11	88.22
13	95.03	91.10	35	93.33	90.00
14	93.77	92.76	36	90.35	86.11
15	94.71	88.93	37	94.17	88.53
16	91.53	90.41	38	95.39	88.18
17	90.82	89.42	39	93.12	90.72
18	92.09	90.69	40	95.71	92.43
19	94.11	91.13	41	93.35	87.09
20	94.58	91.99	42	90.67	88.96
21	92.45	86.25	43	93.52	90.41
22	95.25	88.49	44	93.73	92.64

to the improved $(2+1)$D decomposition process based on the R3D network used in the previous section. The three-dimensional convolution kernel of $3\times3\times3$ is decomposed into a two-dimensional spatial convolution kernel of $1\times3\times3$ and a one-dimensional temporal convolution kernel of $3\times1\times1$. The specific decomposition process is shown in Figure 7, where Conv_s represents the decomposition in the spatial domain and Conv_t represents the decomposition in the time domain.

The structure diagram of the R $(2+1)$ D network model is shown in Figure 8.

4. A Two-Channel Convolutional Network for Calligraphic Style Feature Extraction and Discrimination

The third part constructs a calligraphy style feature extraction and identification system based on a two-channel convolutional network. The authenticity identification process of calligraphy works is generally divided into four steps: scanning of original works, preprocessing of original works, extraction of feature points, comparison and matching of authentic and fake works, and performance evaluation of the identification system. The general process of calligraphy identification is shown in Figure 9(a). The system structure diagram is shown in Figure 9(b).

The computer-aided identification of authenticity of Chinese calligraphy is based on the overall style characteristics of calligraphers, and a database of authentic works of different calligraphers should be built during feature extraction and identification. A feature database is established by extracting feature data of the works. The algorithm first scans the original calligraphy works to obtain digitized works and then divides the pages to obtain single subimages. By extracting the outline skeleton and stroke features of calligraphy characters, the algorithm calculates the feature weight and the authenticity judgment function and constructs the authenticity identification system. Figure 9(b) is the structure diagram of the authenticity identification system.

Figure 10 shows an example of calligraphy image recognition proposed in this paper.

On the basis of the above research, the effect of the calligraphy style feature extraction and identification system based on the dual-channel convolutional neural network proposed in this study is verified, and the results are shown in Table 1.

From the above research, it can be seen that the calligraphy style feature extraction and identification system based on the dual-channel convolutional neural network proposed in this paper has a good performance in calligraphy style feature extraction and identification.

5. Conclusion

The contour of calligraphy is also called edge detection, and the result of extraction is several closed contour curves formed by contour tracking. Contour feature extraction is a common method in image processing. When extracting the features of calligraphy characters, the contour edge features of calligraphy characters are generally extracted first, and the extracted contour features can eliminate a large amount of redundant information, which is beneficial to the acquisition of subsequent feature points. Moreover, the refinement of calligraphic characters is very important in calligraphic character feature extraction and style learning. This paper

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9842467, 1 page
https://doi.org/10.1155/2023/9842467

Retraction

Retracted: Financial and Economic Sequence Forecasting Based on Time Slot Allocation Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wu, "Financial and Economic Sequence Forecasting Based on Time Slot Allocation Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 2340521, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2340521, 13 pages
https://doi.org/10.1155/2022/2340521

WILEY | Hindawi

Research Article

Financial and Economic Sequence Forecasting Based on Time Slot Allocation Algorithm

Yuhan Wu

Henan University of Animal Husbandry and Economy, Zhengzhou 450044, China

Correspondence should be addressed to Yuhan Wu; 81705@hnuahe.edu.cn

Received 20 January 2022; Revised 1 March 2022; Accepted 22 April 2022; Published 16 May 2022

Academic Editor: Chin-Ling Chen

Financial and economic series prediction was very important for the study of economic time series. Because the traditional time series analysis method was mainly suitable for stable data and cannot be applied to unstable economic variables, this paper proposed a knowledge model of financial and economic series prediction time slot algorithm, which provided the basis for solving unstable financial data. Firstly, based on the analysis of relevant financial and economic series prediction methods at home and abroad, this paper applied the time slot allocation algorithm to economic series analysis and prediction and designed a path level time slot allocation algorithm, which introduced a fixed time slot to shorten the time required to complete the path level time slot allocation. Secondly, this paper abstracted the calculation of slot allocation as a convex optimization problem and proposed a round slot allocation algorithm based on constraints and optimal solution. The algorithm can improve the end-to-end throughput and the effective utilization of slots. Finally, the model proposed in this paper was used to predict the indicators of relevant industries, and the simulation results were compared with the standard data. The experimental results showed that the financial and economic series prediction system based on time slot allocation algorithm proposed in this paper can effectively improve the accuracy of financial and economic series prediction, which verified the effectiveness and feasibility of this model.

1. Introduction

Although the research on structural mutation in time series analysis has not been long, the research problems in this field have been widely concerned by relevant scholars. Moreover, the analysis of time series with structural mutation is still the frontier and hot topic in this field. The research shows that the trend stationary process with structural mutation has similar characteristics with the unit root process, which is easy to be confused. Therefore, many unit root test methods are prone to misjudgment when distinguishing these two types of processes, which can easily lead to the decline of test ability and the distortion of test level [1]. When considering structural mutation, a new unit root test method is needed to determine the stationarity of the sequence data generation process. At present, researchers in econometrics have proposed several unit root test methods with structural mutation [2]. However, due to the different research basis or relevant assumptions of different methods, using different test methods for the same problem in empirical analysis may lead to different or even contradictory conclusions. Therefore, it is an urgent problem to put forward an objective and scientific unit root test method in theory, which has very important value and significance both in theory and in practical application [3].

The purpose of econometric theory research is to help people understand and grasp the law of economic development more objectively and scientifically. Relevant theories and new methods can make up for and correct the shortcomings of existing methods and then make the conclusion of empirical analysis more scientific. Through the organic combination of unit root test of time series and structural catastrophe theory, we can not only improve and develop relevant theories, but also analyze and solve practical problems. The unit root test of structural mutation is not only the theoretical basis of econometric method and time series analysis, but also the premise of studying the relationship between economic variables such as cointegration analysis and Granger causality test. In addition, because panel data has two-dimensional characteristics of space and time, unit root test is also of great significance to panel data analysis, such as the stationarity analysis of time dimension. In view of the impact of structural mutation on unit root test, this paper proposes a unit root test method and program including structural mutation, which will help to improve the effectiveness of traditional unit root test and provide a scientific basis for revising and perfecting the theoretical research of unit root test.

For the research of practical economic problems, the stability of data generation process can be applied to the effectiveness analysis of economic policies. For example, when an economic variable is proved to be produced by a stable process of structural mutation, it indicates that the economic variable has changed the original growth path at a certain time due to the influence of various factors. When the government wants to change its long-term development trend, it can achieve the purpose of regulation and control by implementing corresponding economic policies. It can be seen that the research on structural mutation and unit root test has certain economic and practical significance.

For the observation of time series, due to the joint action of many factors, there may be a few observations that are significantly different from other observations, which are called outliers. For the whole time series, although the number of outliers is small, their existence may have a certain impact on model identification, parameter estimation, model diagnosis, and even prediction. Some outliers may contain very valuable information. For example, the outliers of the birth rate may reflect changes in national policies and social development, and the outliers of the stock market may reflect changes in national economic conditions. Therefore, the study of outliers in time series has extremely important theoretical and practical significance.

2. Related Works

As the research basis of time series analysis, unit root test is mainly used to analyze the stationarity of time series. The research shows that if the growth path of the economic sequence is affected by the deterministic trend function, the external random impact will cause a temporary deviation from the trend term of the sequence without changing the growth path of the economic sequence. Since this economic series is stable after removing the trend item, the economic variable belongs to the trend stable type or the determined trend type. If there is a unit root in the data generation process of time series, any external random shock may change the growth path of economic series. At this time, this economic series belongs to nonstationary type, which is also called random trend process or unit root process.

In fact, the unit root test of structural mutation proposed earlier is to add the dummy variable reflecting structural mutation (i.e., the mutant dummy variable) to the unit root test of DF (ADF) [4]. Although this method gives the asymptotic distribution and asymptotic critical value of the unit root of the test statistics, the hypothesis of external source mutation point has been questioned. The endogenous determination method of mutation point mainly uses the relationship between data to determine the location of mutation point. Among them, common methods, for example, estimate the mutation point according to the minimum value of the proposed unit root test statistic [5] and estimate the mutation point according to the maximum value of the test statistic (absolute value of F statistic or F statistic) according to the significance of the coefficient estimator of mutation virtual variable [6]. The mutation point is tested by the unit root of DF (ADF) statistic of mutation dummy variable regression [7]. According to the original hypothesis, there are generally two types of testing for mutation points. One is that the original hypothesis has no structural mutation, and the other is that the original hypothesis allows structural mutation [8].

Some people have analyzed the influence of the wrong configuration of mutation points on the asymptotic distribution of unit root test statistics and the properties of finite samples under horizontal mutation [9, 10]. Through the simulation test, it is found that when there is a structural mutation in the process of real data generation, the mutation point is estimated based on the minimum value of the unit root test statistic or the maximum value of the significance test statistic of the mutation pseudovariable coefficient [11, 12]. It is often one cycle earlier than the real mutation point, which will lead to the horizontal distortion of the unit root test and reduce the test power. Some people determine the mutation point according to the minimum value of BIC statistics of test regression or take the time corresponding to the minimum value of the sum of squares of correlation regression residuals as the estimated value of the mutation point [13]. In addition, some scholars have studied the unit root test statistics of endogenous mutation points and the unit root test statistics of exogenous mutation points under different mutation models according to the speed of convergence from the mutation point estimated by the minimum sum of squares of residuals to the true mutation point [14].

After introducing structural mutation into unit root test, the methods related to unit root test have been widely used in structural mutation [15]. For example, some people analyzed the change trend of mutation point and the properties of mutation LM unit root test statistics under the framework of AO model [16]. Some scholars believe that the mutation exists under the assumption and study the properties of LM unit root test statistics to determine the mutation point according to the minimum value of LM statistics [17]. Later, some scholars further studied the LM unit root test statistics of structural mutation. The maximum or minimum value of statistics is tested according to the significance of mutation virtual variable coefficient, and the minimum value of ADF statistics is obtained by using the unit root test of structural mutation, so as to determine the minimum value of LM statistics at the mutation point [18]. In addition, some people introduce structural mutation into KPSS test, construct the corresponding test statistics, and study its asymptotic distribution and finite sample properties [19]. It is known from the existing research that the traditional unit root test method generally believes that the data generation process of time series is linear, and if it is applied to the unit root test of nonlinear trend, it may produce wrong results. Therefore, the single root test with nonlinear trend structure has certain practical significance.

3. Path Level Time Slot Allocation Algorithm and Knowledge Model

3.1. Calculation Method of Path Slot Allocation. Considering the shortcomings of the traditional system in analyzing a large amount of economic data, in order to optimize the economic data analysis nodes, this paper mainly uses the intelligent model to process the economic data when predicting the economic sequence. It is assumed that the information about the available time slots of all nodes on the path has been obtained, and the time slots are allocated to all nodes on the path in order to maximize the end-to-end throughput. Because the bottleneck node on the service path limits the end-to-end throughput to a certain extent, the maximum number of time slots can be obtained at the concave point through time slots. Therefore, the path level slot allocation problem can be abstracted as a convex optimization problem and solved with the help of computational tools.

The goal of path slot allocation calculation is to maximize the number of time slots occupied by time slot pits on the path, and time slot pits are the nodes that obtain the least time slots in the time slot allocation. The specific problems are described as follows:

$x_{i,j}$ indicates whether node i occupies time slot j in this time slot allocation process, $x_{i,j} \in \{0,1\}, x_{i,j} = 1$ shows that node i occupies time slot j, and $x_{i,j} = 0$ indicates that it does not occupy. Among them, i indicates the serial number of node on the path, j is the number of data time slot in the compound frame, we set the number of nodes on the path as n and the total number of time slots as m, and then there are $1 \le j \le m, 1 \le i \le n$. Therefore, the number of time slots occupied by time slot concave points can be expressed as follows:

$$N_{\min} = \min_{i\in[1,n]} \sum_{j=1}^{m} X_{i,j}, \tag{1}$$

where $N_{\min}$ represents the number of time slots occupied by the slot pits, and the goal problem of our time slot allocation is to maximize the value of $N_{\min}$; that is, $\max(N_{\min})$ is expressed as follows:

$$\max(N_{\min}) = \max\left(\min_{i\in[1,n]} \sum_{j=1}^{m} X_{i,j}\right), \tag{2}$$

where $X_i = [X_{i,1}, X_{i,2}, X_{i,3}, \ldots, X_{i,m}]^T$, $X = [X_{1,1}, X_{1,2}, \ldots, X_{1,m}, X_{2,1}, X_{2,2}, \ldots, X_{2,m}, \ldots, X_{i,\ j}, \ldots, X_{n,m}]^T$,

$$\sum_{j=1}^{m} X_{i,j} = [1,1,1,\ldots,1] \cdot X_i. \tag{3}$$

Node i obtains the number of time slots $S = \sum_{j-1}^{m} X_{i,j}$ in this time slot allocation, and it has the following form [20]:

$$S = Q_i \cdot X, \tag{4}$$

where $Q = [0,0,\ldots,0,1,1,\ldots,1.0,0,\ldots,0]$, which is a vector of $n \cdot m$, preceded by $(i-1)\cdot m$ consecutive 0, followed by m consecutive 1, followed by all 0. Then the problem P_0 is of the following form:

$$P_0 = \max\left(\min_{i\in[1,n]} (Q \cdot X)\right), \quad i \in [1,n]. \tag{5}$$

Formula (5) is our original problem P. In order to solve it, we introduce the variable W, which is as follows:

$$W = \operatorname*{mim}_{i\hat{\text{I}}[1,n]} (Q \cdot X), \quad i \in [1,n]. \tag{6}$$

In this way, our problem P_0 is transformed into

$$P_1 = \max(w). \tag{7}$$

In problem P_1, w must satisfy

$$w \le Q_i \cdot X, \forall i \in [1,n]. \tag{8}$$

Then the original problem P_0 is transformed into P_1 in formula (9). To solve P_1, X and w must be constrained. We might as well set $V = [X^T, w]^T$, that is, $V = [X_{1,1}, X_{1,2}, \ldots, X_{1,m}, X_{2,1}, X_{2,2}, \ldots, X_{2,m}, \ldots, X_{i,j}, \ldots, X_{n,m}]^T$, and we only need to constrain V. Then the problem P_1 can be described as [21]

$$P_1 = \max([0,0,\ldots 0,1] \cdot V). \tag{9}$$

Condition 1 is described as $w \le Q_i \cdot X, \quad \forall i \in [1,n]$. The constraints in formula (8) can be reduced to

$$w \le A_i \cdot V, \quad i \in [1,n], \tag{10}$$

where $A_i = [Q_i, 0], w = [0,0,\ldots,0,1] \cdot V$; then the following expression is satisfied:

$$([0,0,\ldots,0,1] - A_i) \cdot V \le 0, \quad \forall i \in [1,n]. \tag{11}$$

It can also be written in the following form:

$$[-Q_i, 1] \cdot V \le 0, \quad \forall i \in [1,n]. \tag{12}$$

The following expression can be obtained:

$$Q \cdot V \le [0,0,\ldots 0]^T, \tag{13}$$

where Q is a matrix of $n \cdot (n \cdot m + 1)$ in the following form:

$$Q = \begin{bmatrix} -Q_{1,1} \\ -Q_{2,1} \\ \vdots \\ -Q_{n,1} \end{bmatrix}. \tag{14}$$

Formula (13) is the first constraint condition of problem P_1.

Condition 2 is described as follows: any node i on the path has its own set of occupiable time slots, and the time slot finally obtained in this time slot allocation must belong to its own set of occupiable time slots.

We use $X_i = [X_{i,1}, X_{i,2}, X_{i,3}, \ldots\ldots X_{i,m}]^T$ to represent the time slot allocation result of node i. If i occupies time slot j, then the corresponding X can be set to 1. We introduce $U_i = [u_{i,1}, u_{i,2}, u_{i,3}, \ldots, u_{i,m}]^T$ about the set of occupied time slots of node i, where $u_{ij} \in \{0, 1\}, U_i$ is the input parameter of the target problem. In this way, the constraints of node i can be expressed as

$$U_i \cdot X_i = 0, \quad i \in [1, n]. \tag{15}$$

From formula (15), when $u_{ij} = 1$, $X_{i,j}$ must be 0. At this time, node i cannot occupy time slot j. When $u_{ij} = 0$, $X_{i,j} \in \{0, 1\}$. By analogy, each node on the path meets the following conditions:

$$\begin{aligned} U_1 \cdot X_1 &= 0, \\ U_2 \cdot X_2 &= 0, \\ U_3 \cdot X_3 &= 0, \\ &\vdots, \\ U_n \cdot X_n &= 0. \end{aligned} \tag{16}$$

Through sorting, the following expression can be obtained:

$$U \cdot V = [0, 0, \ldots, 0]^T, \tag{17}$$

where U is an $n \cdot (n \cdot m + 1)$ matrix as shown in the following formula:

$$U = \begin{pmatrix} U_1 & O & O & O & \cdots & O & 0 \\ O & U_2 & O & O & \cdots & O & 0 \\ O & O & U_3 & O & \cdots & O & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O & \cdots & O & O & U_{n-1} & O & 0 \\ O & \cdots & O & O & O & U_n & 0 \end{pmatrix}, \tag{18}$$

where O is an m-dimensional vector with all zero elements. In summary, formula (17) is the second constraint condition of the objective problem.

Condition 3 is described as follows: the node time slot occupancy within two hops does not conflict, which can be expressed as follows:

When $i = 1$, $x_{ij} + x_{i+1,j} + x_{i+2,j} \le 1, \quad 1 \le j \le m$.

When $i = 2$, $x_{i-1,j} + x_{i,j} + x_{i+1,+j} + x_{i+2,+j}, \quad 1 \le j \le m$.

When $3 \le i \le n-2$, $x_{i-2,j} + x_{i-1,j} + x_i + x_{i,j} + x_{i+1,+j} + x_{i+2,+j} \le 1, \quad 1 \le j \le m$.

When $i = n-1$, $x_{i-1,j} + x_{i-2,j} + x_{i,j} + x_{i+1,+j} \le 1, \quad 1 \le j \le m$.

When $i = n$, $x_{i-1,j} + x_{i-2,j} + x_{i,j} \le 1, \quad 1 \le j \le m$.

Each node on the path should have a restriction condition for all time slots. For any node i, m restriction conditions are required to ensure that the time slot occupancy within two hops does not conflict. Taking the nodes satisfying $3 \le i \le n-2$ as an example, there is the following expression:

$$C_i \cdot V \le [1, 1, \cdots, 1]^T. \tag{19}$$

Node i requires m restriction conditions, and C_i is a matrix of $m \cdot (n \cdot m + 1)$. According to each restriction condition of node i, the specific form of C_i is as follows:

$$C_i = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}. \tag{20}$$

The elements with the value of 1 in the first row of the matrix in formula (20) correspond to $x_{i-2,1}$, $x_{i-1,1}$, $x_{i,1}$, $x_{i+1,1}$, and $x_{i+2,1}$ in V, respectively, which are the constraint coefficients of node i for time slot 1. The elements with the value of 1 in the second row correspond to $x_{i-2,j}$, $x_{i-1,j}$, $x_{i,j}$, $x_{i+1,j}$, and $x_{i+2,j}$ at the center, respectively, which are the constraint coefficients of node i for time slot 2. The elements with the value of 1 in the j-th row correspond to $x_{i-2,j}$, $x_{i-1,j}$, $x_{i,j}$, $x_{i+1,j}$, and $x_{i+2,j}$, respectively, which are the constraint coefficients of node i for time slot j. It can be seen that the matrix C_i is the constraint coefficient matrix of node i for all time slots, that is, the coefficient C_i when $i = 1$, $i = 2$, $i = n-1$, and $i = n$. Similarly, the following formula can be derived:

$$C \cdot V \le [1, 1, \ldots, 1]^T. \tag{21}$$

There is a total of $n \cdot m$ constraint conditions for n nodes. Therefore, C in formula (21) is a matrix of $(n \cdot m) \cdot (n \cdot m + 1)$, as shown below:

$$C = [C_1, C_2, \ldots, C_n]^T. \tag{22}$$

In summary, formula (21) is the third constraint condition of the target problem P_1.

Condition 4 is described as:$X_{i,j} \in \{0,1\}$, and the value of $Q_{1,1}$ can only be 0 and 1, which is equivalent to $X_{i,j}(X_{i,j}-1)=0$. It can be transformed into the form of $V^T Q_{i,j} V - P_{i,j} V = 0$; that is, construct $Q_{i,j}$ and $Q_{i,j}$.

$$\begin{gathered} V^T Q_{i,j} V - P_{i,j} V = 0, \\ \Downarrow, \\ x_{i,j}^2 - x_{i,j} = 0. \end{gathered} \tag{23}$$

We start from X to construct Q and P and get the following expression:

$$Q_{1,1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{24}$$

$$P_{1,1} = [1, 0, 0, \ldots, 0]. \tag{25}$$

As shown in formula (24), $Q_{1,1}$ is a matrix of $(n \cdot m + 1) \cdot (n \cdot m + 1)$, as shown in formula (25). $P_{1,1}$ is a $P_{1,1}$-dimensional vector.

Similarly, $Q_{i,j}$ can be obtained. For $\forall j \in [1,m]$ and $\forall i \in [1,n]$, then $Q_{i,j}$ is a matrix of $(n \cdot m + 1) \cdot (n \cdot m + 1)$, where $Q_{i,j}(i \cdot j - 1, i \cdot j - 1) = 1$, and the values of other elements are all 0, and $Q_{i,j}^T = Q_{i,j}$.

It is easy to know that $Q_{i,j}$ is a $n \cdot m + 1$-dimensional vector, where the value of the $(i-1) + j$-th element is 1, and the other elements are all 0.

In summary, condition 4 can be transformed into $m \cdot n$ constraint conditions, which satisfy the following relationship:

$$\begin{gathered} V^T Q_{1,1} V - P_{1,1} V = 0, \\ V^T Q_{2,2} V - P_{2,2} V = 0, \\ \vdots, \\ V^T Q_{n,m} V - P_{n,m} V = 0. \end{gathered} \tag{26}$$

The above constraint conditions include quadratic form $V^T Q_{i,j} V$, so we might as well set $f_{ij}(V) = V^T Q_{i,j} V$. Assuming that V_0 is any feasible point in the feasible region, we carry out a first-order Taylor expansion of $f_{ij}(V)$ at $V = V_0$ and obtain the following relationship:

$$f_{ij}(V) = f_{ij}(V_0) + f_{ij}(V_0)(V - V_0), \tag{27}$$

$$f_{ij}(V) = V_0^T Q_{i,j} V_0 + 2Q_{i,j} \cdot V_0 (V - V_0). \tag{28}$$

$V^T Q_{i,j} V - P_{i,j} V = 0$ is equivalent to

$$V^T Q_{i,j} V - P_{i,j} V \le 0, \tag{29}$$

$$P_{i,j} V - V^T Q_{i,j} V \le 0. \tag{30}$$

Substituting formula (27) into formula (30), the constraints of formula (29) and formula (30) are equivalent to

$$\begin{gathered} 0 \le x_{i,j} \le 1 \quad \forall j \in [1,m], \forall i \in [1,n], \\ P_{i,j} V - \left(V_0^T Q_{i,j} V_0 + 2Q_{i,j} \cdot V_0 (V - V_0)\right) \le 0, \quad \forall j \in [1,m], \forall i \in [1,n]. \end{gathered} \tag{31}$$

Formula (31) is the fourth constraint.

In summary, the target problem P_1 can be transformed into problem P_2, which can be expressed as follows:

$$\begin{gathered} P_2 = \max([0, 0, \ldots, 0, 1] \cdot V), \\ Q \cdot V \le [0, 0, \ldots, 0]^T, \\ U \cdot V \le [0, 0, \ldots, 0]^T, \\ C \cdot V \le [0, 0, \ldots, 0]^T, \\ 0 \le x_{i,j} \le 1, P_{i,j} V - \left(V_0^T Q_{i,j} V_0 + 2Q_{i,j} \cdot V_0 (V - V_0)\right) \le 0 \forall j \in [1,m], \forall i \in [1,n]. \end{gathered} \tag{32}$$

3.2. Parameter Analysis of Time Slot Allocation Algorithm. So far, we have described the objective problem P as a convex optimization problem, and we only need to organize the matrix U according to the initial set of occupied time slots of each node on the path; that is, we can calculate V with the help of CVX tool in MATLAB, where X^T is the situation of time slots obtained by each node in this assignment, and W is the optimal solution of the objective problem, that is, the number of time slots obtained by the time slot concave point. When the obtained W value is greater than the number of time slots required by the nodes, it indicates that the current time slots are sufficient and only the required number of time slots needs to be taken from the time slots allocated to each node.

From the above analysis, the time required for the time slot allocation algorithm based on USAP protocol to complete the path time slot allocation can be expressed as follows:

$$t_{to} = n(T_s + T_c), \tag{33}$$

where T_s indicates the superframe time and T_c is the multiframe time. According to the solution idea design plan, then this time can be expressed by the following expression:

$$t_{\text{to}} = t_{\text{cal}} + t_{\text{sum}} + t_{\text{dis}} + t_{\text{res}}, \tag{34}$$

where t_{sum} and t_{dis} are the time required for the time slot aggregation and the calculation of the result distribution, respectively. If the classical USAP protocol superframe structure as described is used, the aggregation of time slots and distribution of results can only be encapsulated in NMOP broadcast out at the broadcast time slot. The service flow path in a self-organizing network is random, and it is basically impossible to be in the same direction as the occupation of broadcast time slots by nodes on the service flow path. If the BS occupancy happens to be in the opposite direction of the service flow transmission, it is assumed that there are n nodes on the path, and then $t_{\text{sum}} = nT$, and by the same token $t_{\text{dis}} = nT$, this overhead is not tolerable. Therefore, we need to let each node occupy a small number of time slots upon entering the network to transmit the relevant control frames for time slot allocation. Therefore, PL-TDMA introduces a certain number of fixed time slots into each complex frame of superframe for transmitting control frames related to time slot allocation. At this time, nodes occupy fixed time slots without reservation and conflict. The fixed time slots only transmit control frames, and their length should be much smaller than that of data time slots so that their overhead can be neglected. Thus, PL-TDMA can reduce the length of t_{sum} and t_{dis} from superframe multiples to complex frame multiples.

In formula (34), t_{cal} is the same as the USAP protocol allocation algorithm, and t_{sum} is related to the number of hops n. According to the PL-TDMA frame format design, nodes occupy a fixed time slot for at least 4 compound frames, the number of nodes on the path is n, the length of a compound frame is T_c, the length of a superframe is T_s, $t_{\text{sum}} = n \cdot 4T_c$, and $t_{\text{dis}} = t_{\text{sum}} \cdot 4T_c$. The reservation time is the same as the classical USAP protocol for one superframe, that is, $t_{\text{res}} = T_s$. By bringing t_{sum}, t_{dis}, t_{cal}, and t_{res} into formula (34), the following expression can be obtained:

$$t_{to} = (8n + 1) \cdot T_C + T_S. \tag{35}$$

From formula (33) and formula (35), it can be seen that the time required for PL-TDMA path time slot allocation is significantly reduced compared to USAP protocol when the hop count is greater than 2.

PL-TDMA classifies the time slots in the superframe into three types, which are synchronous broadcast time slots, fixed time slots and dynamic time slots. Among them, synchronous broadcast time slot is the time slot used by nodes to broadcast service frames, which is similar to NMOP information frames in USAP and is used for information interaction between nodes to complete network establishment and maintenance. Fixed time slots are mainly used to transmit control frames related to time slot reservation, while dynamic time slots are used for service data transmission. Assuming that the maximum number of nodes in the network is $4n$, the super frame structure is shown in Figure 1.

The PL-TDMA frame structure is cyclic in a unit cycle of superframes. Each superframe consists of $4n$ complex frames, each of which is composed of one simulcast time slot, n fixed time slots, and m dynamic time slots. A node is connected to the network through distributed information interaction with other nodes, while the node occupies one synchronous broadcast time slot in each superframe. After the node is successfully connected to the network, it occupies the fixed time slots through the mapping of broadcast time slots. Nodes do not occupy dynamic time slots directly after joining the network, but use them as reserved time slot resources. When the service load of the node reaches the dynamic time slot occupancy condition, it can occupy the free dynamic time slot resource through the dynamic time slot allocation algorithm.

The synchronous broadcast time slot is the control time slot of the protocol. In the synchronous broadcast time slot, the node broadcasts the service frame containing the information of itself and neighboring nodes. The establishment and maintenance of the network are completed by the interaction of the service frame between the nodes.

The introduction of fixed time slots allows nodes to transmit information about dynamic time slot allocation after they are on the network, which will greatly reduce the time required for the completion of path time slot allocation. The allocation of fixed time slots is based on the mapping rules and the broadcast time slots occupied by the nodes. The node occupies its own broadcast time slot when it enters the network according to the node's occupancy of the broadcast time slot within two hops, so it can be guaranteed that the occupancy of the broadcast time slot is nonconflicting. Therefore, it is easy to generalize the mapping rules to ensure that the fixed time slot occupation is also conflict-free in the two-hop range. In order to improve the time slot utilization, the fixed time slot occupation is determined by the number of nodes in the two-hop range.

According to the superframe structure, the maximum number of nodes in the two-hop range in the network is set to $4n$, and then each superframe contains $4n$ broadcast time slots, and each multiframe contains n fixed time slots, and the time slot number is from FS_0 to FS_{4n-1}, cycling in sequence with a period of four multiframes. The fixed time

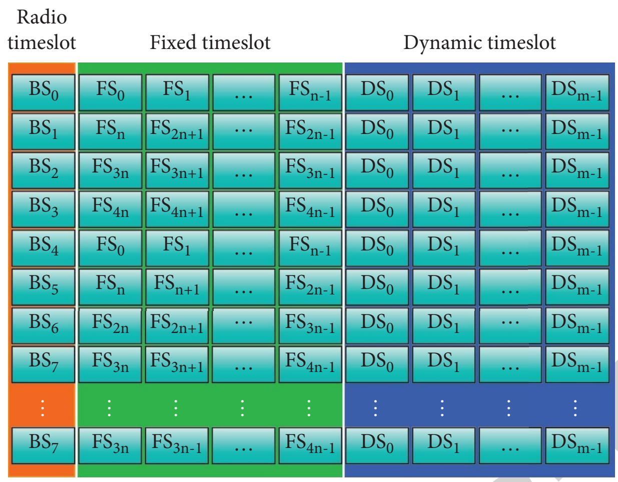


Figure 1: PL-TDMA super frame structure diagram.

slots are occupied according to the number of nodes in the two-hop range, and the specific occupation method is as follows:

(1) When the number of nodes in the two-hop range does not exceed n, the node can occupy a fixed time slot in each multiframe. We assume that the broadcast time slot occupied by the node is the BS, and the occupied set mapped to the fixed time slot is $S = \{FS_i + FS_{n+i} + FS_{2n+i}, FS_{3n+i}\}$; that is, each multiframe occupies a fixed time slot.

(2) When the number of nodes in the two-hop range is greater than n and not more than $2n$, the node can occupy a fixed time slot in every two multiframes. We assume that the broadcast time slot occupied by the node is BS, and the occupied set mapped to the fixed time slot is $S = \{FS_i, FS_{2n+i}\}$; that is, every 2 multiframes occupy a fixed time slot.

(3) When the number of nodes in the two-hop range is greater than $2n$ and not more than $4n$, the node can occupy a fixed time slot in every four multiframes. We assume that the broadcast time slot occupied by the node is BS, and the occupied set mapped to the fixed time slot is $S = \{FS\}$; that is, every 4 multiframes occupy a fixed time slot.

The abovementioned fixed time slot allocation method makes the node occupy a certain fixed time slot after entering the network. The occupancy of the fixed time slot provides the transmission condition of the relevant protocol control frame for the dynamic time slot allocation and is the key to shorten the path delay of the time slot allocation by the PL-TDMA designed in this paper. The node first transmits the protocol data frame related to the dynamic time slot allocation in the fixed time slot and then can transmit other data frames.

According to the above solution and the abstraction of the path slot allocation problem, this paper designs a path level slot allocation scheme, as shown in Figure 2(a). The computing node of the service flow path summarizes the time slot information that can be occupied by each node on the path and performs time slot allocation calculation. The calculation method is based on the round distribution, and the distribution principle is the following: within two hops, the node time slot occupation does not conflict and makes full use of the space multiplexing characteristics of the time slot to ensure that each node on the path gets the same number of time slots after the allocation is completed. The main steps of the PL-TDMA time slot allocation algorithm are three steps, as shown in Figure 2(b).

Service flow detection is the first step of the algorithm. After the MAC layer detects the arrival of a new service flow from the application layer, it determines the service flow type based on the transport layer header information of the service data. The source and destination addresses are obtained based on the self-organizing network frame header information, and the routing table of the self-organizing network is queried to obtain the information of all nodes on the path. We obtain the service flow rate based on the rate at which the service flow arrives at the MAC layer and define a self-increment as the service flow number. The above service flow information is stored and maintained locally and encapsulated into a service flow information frame with the service flow information format shown in Figure 3(a). The time slots available for nodes are appended to the end of the service information frame in ascending order by time slot number, and the format is shown in Figure 3(b). The destination address and receiving address in the header information of the data frame of the self-organized network are filled as the next hop of the path, and the next hop node receives the frame, saves the service flow information locally, and appends its own occupiable time slot information at the end of the frame to the path to the next hop node, until the computing node, which is the previous hop node of the destination node. The format of the information received by the computing node is shown in Figure 3(c).

We assume that there are M nodes on the path. After the above steps, each node on the path can obtain the type of service flow, the arrival rate of service flow V_{in}, and the set of nodes on the service path $S = \{n_1, n_2, \ldots, n_M\}$. The order of the nodes in the set is arranged in the order on the path.

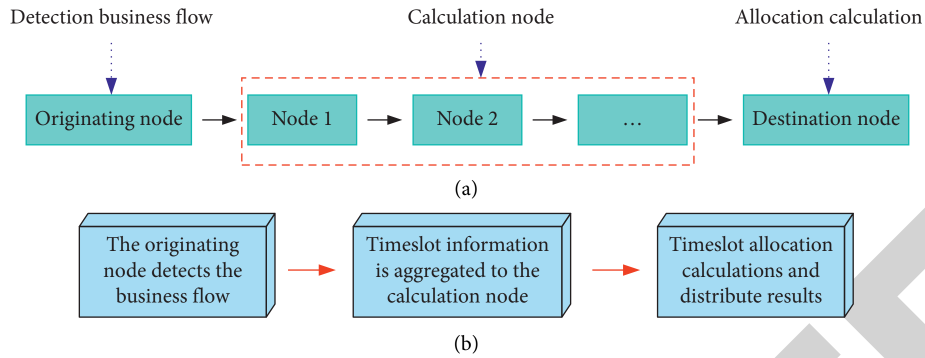


Figure 2: The main steps of PL-TDMA time slot allocation algorithm. (a) Schematic diagram of path-level time slot allocation. (b) Schematic diagram of PL-TDMA time slot allocation algorithm steps.

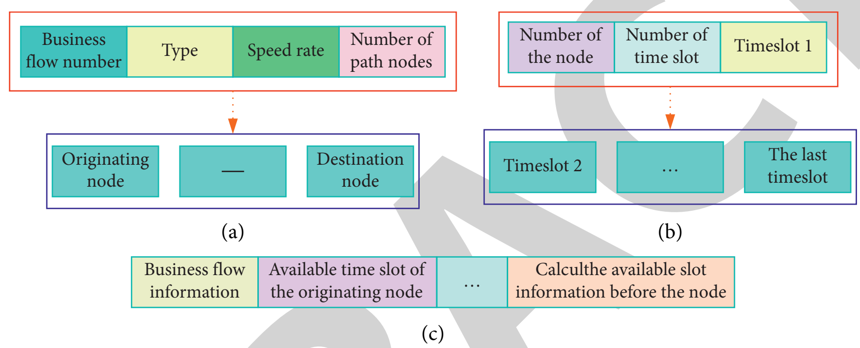


Figure 3: Summary of message format. (a) Business flow information format. (b) Node available time slot information format. (c) Format of summary message information.

In addition to service flow information, computing nodes can also obtain the set of time slots that can be occupied by each node on the path.

3.3. Prediction Model of Financial and Economic Series Based on Time Slot Allocation Algorithm. The subsequent prediction of error in this paper is based on the autocorrelation of error series. Generally, modeling within the scope of econometrics assumes that there is no correlation between the early and late stages of random errors. However, in an economic system, there may be correlation between the front and rear economic variables, so the random error can not meet the assumption of no autocorrelation. For autocorrelation test, it mainly includes graphic test method and DW test method. Among them, the graphical method is relatively intuitive. It uses the least square estimation method to obtain its parameters and then uses the estimation method to draw the residual scatter diagram after obtaining the residual term.

From the model established in this paper to predict the change trend of the original sequence, it is known that, due to the defects of the model itself, the predicted residuals may have a certain autocorrelation. Therefore, a model can be established to predict the residual term and correct the prediction results. As shown in Figure 4, it is a process of predicting the residual term and correcting the result by using the model.

Taking regional economy as the research object, this paper constructs the architecture of regional economic trend prediction system. The design of the architecture is mainly based on the general process of data mining and adopts the construction method of conventional application platform, as shown in Figure 5.

As shown in Figure 6, it shows the functional structure of the regional economic trend prediction and analysis system.

The regional economic trend prediction and analysis module mainly uses the constructed regional economic trend prediction model to predict the change trend of relevant indicators in regional economic analysis. The prediction indicators mainly include the development prediction of market subjects, the development prediction of industrial structure, and the development prediction of enterprises. As shown in Figure 7, it describes the process of predicting the future development trend of regional economy from macro- to microaspects according to the main body of regional economy market, industrial structure, and enterprise development.

4. Experimental Analysis and Discussion

In order to test the financial and economic series prediction model based on time slot allocation algorithm proposed in this paper, the network simulation platform is used to simulate the model in various network scenarios. At present,

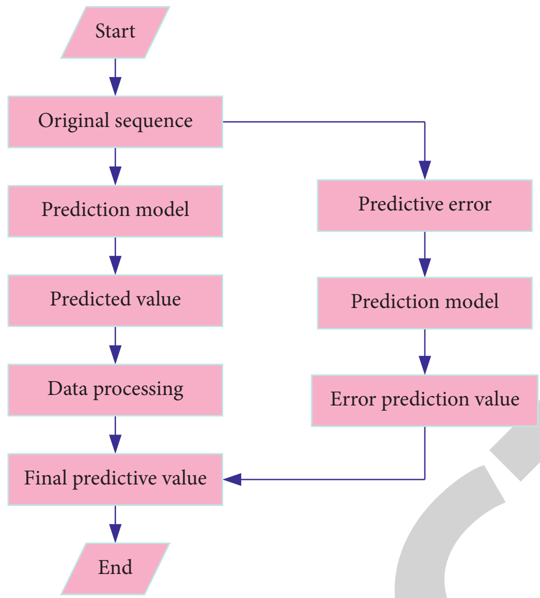


FIGURE 4: The process of using the model to predict the remaining items and correct the results.

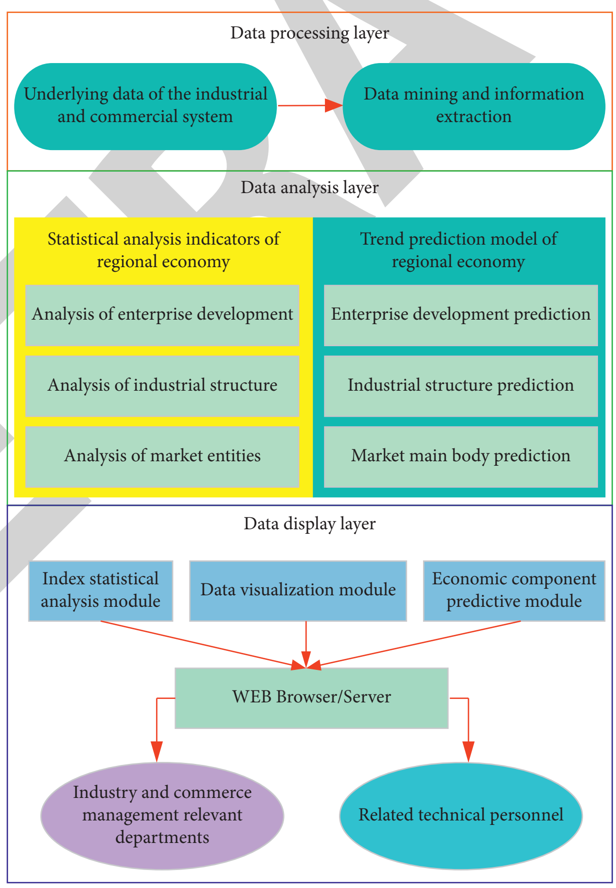


FIGURE 5: Architecture construction process of regional economic trend prediction system.

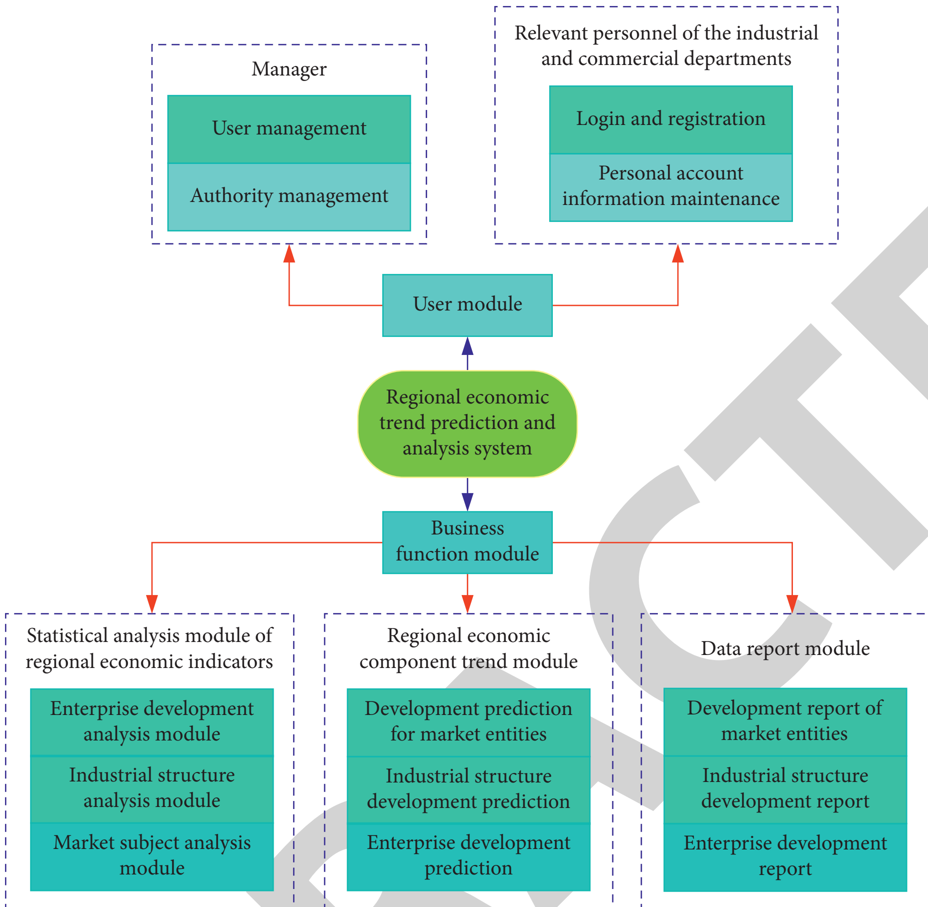


Figure 6: Composition diagram of regional economic trend prediction and analysis system.

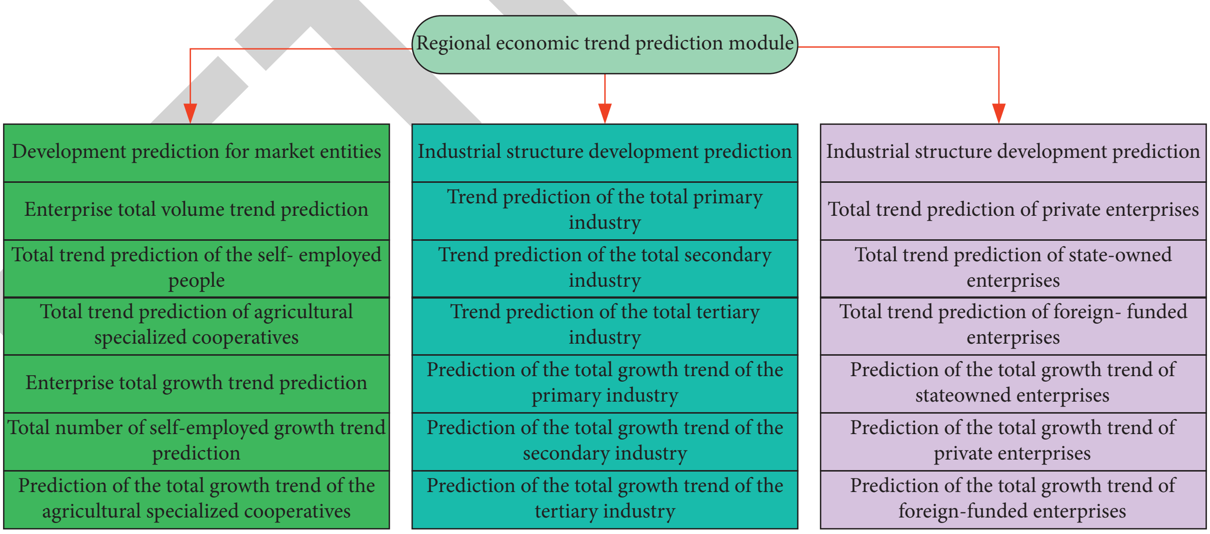


Figure 7: Composition diagram of data report export module.

the mainstream simulation tools include OPNET, NS, and MATLAB. In this paper, OPNET is used as a simulation tool to simulate the performance of the proposed BR-TDMA time slot allocation algorithm.

Through the financial and economic series prediction system based on time slot allocation algorithm, the relevant models in this paper are verified. This paper takes the China Purchasing Managers' Index (PMI) as the economic prediction index and the PMI index from January 2008 to December 2020 as the benchmark (data source: National Bureau of Statistics). In the experiment, the model proposed in this paper is used to predict the indicators of relevant

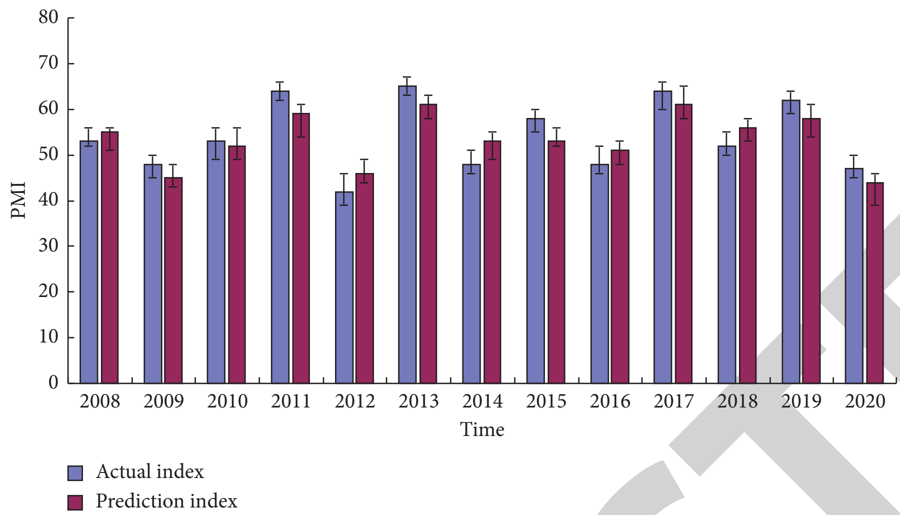


Figure 8: Comparison diagram of manufacturing PMI.

Figure 9: Comparison diagram of nonmanufacturing PMI.

industries, and the simulation results are compared with the standard data. As shown in Figure 8, the comparison results of manufacturing PMI index are shown. As shown in Figure 9, it reflects the comparison results of PMI index of nonmanufacturing industry.

In addition, in order to further verify the effectiveness of the model in this paper, based on the above experimental verification, the model proposed in this paper and the system model in literature [18] are used to predict the statistical economic series. As shown in Table 1, it reflects the prediction results of economic series obtained by using different models.

From the comparison between the model prediction results and the actual values reflected in Figures 8 and 9, it can be seen that the algorithm performance obtained by using the model in this paper is basically consistent with the actual results. From the prediction results of different models on economic series reflected in Table 1, it can be seen that the financial and economic series prediction model based on time slot allocation algorithm constructed in this paper has good robustness and effectiveness, indicating that the model can provide a certain theoretical basis for predicting financial and economic series.

Table 1: Prediction results of economic series obtained by using different models.

Num.	The method of this paper	The method of [18]
1	85.10	70.15
2	85.29	63.88
3	81.01	64.11
4	82.40	67.74
5	86.48	61.40
6	83.84	59.21
7	82.58	59.11
8	83.84	62.91
9	82.32	65.52
10	83.23	62.12
11	81.61	64.25
12	83.71	60.81
13	86.25	68.71
14	82.89	61.61
15	86.37	58.55
16	83.49	68.94
17	83.16	68.15
18	85.50	69.98
19	84.56	67.88
20	84.12	60.84
21	83.78	63.11
22	82.80	61.22
23	86.29	70.45
24	86.05	70.13
25	82.28	61.09
26	81.39	67.39
27	81.07	61.65
28	86.33	62.70
29	83.51	65.50
30	83.44	68.31
31	84.84	67.26
32	81.56	62.55
33	81.72	69.53
34	86.85	69.61
35	85.06	60.02
36	86.49	67.07
37	84.79	61.96
38	83.98	67.94
39	84.53	66.66
40	81.27	65.79

5. Conclusion

The traditional time series analysis method was difficult to apply to the unstable economic variables in the financial industry. Therefore, in order to effectively solve the analysis of unstable financial data, this paper proposed a knowledge model of time slot algorithm for financial and economic series prediction. In order to effectively analyze and predict the economic sequence, this paper designed a path level slot allocation algorithm based on the slot allocation algorithm, and used the fixed slot to improve the efficiency of path level slot allocation. Then, a round slot allocation algorithm based on constraints and optimal solution was proposed, which can improve the end-to-end throughput and the effective utilization of time slots. Finally, this paper forecast and analyzed the relevant industry indicators through experiments and using the algorithm. The results showed that the financial and economic series prediction model based on time slot allocation algorithm proposed in this paper can better realize the effective prediction of financial and economic series. The algorithm proposed in this paper had important reference significance for the application of computer intelligent algorithm in the financial field.

Data Availability

The labeled dataset used to support the findings of this study is available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was sponsored by the Henan University of Animal Husbandry and Economy.

References

[1] W. Montford and R. E. Goldsmith, “How gender and financial self-efficacy influence investment risk taking,” *International Journal of Consumer Studies*, vol. 40, no. 1, pp. 101–106, 2016.
[2] S. M. Wanjohi, J. G. Wanjohi, and J. M. Ndambiri, “The effect of financial risk management on the financial performance of commercial banks in Kenya,” *International Journal of Finance and Banking Research*, vol. 3, no. 5, pp. 70–81, 2017.
[3] J. Magendans, J. M. Gutteling, and S. Zebel, “Psychological determinants of financial buffer saving: The influence of financial risk tolerance and regulatory focus,” *Journal of Risk Research*, vol. 20, no. 8, pp. 1076–1093, 2017.
[4] C. Lucarelli, P. Uberti, and G. Brighetti, “Misclassifications in financial risk tolerance,” *Journal of Risk Research*, vol. 18, no. 4, pp. 467–482, 2015.
[5] E. Y. Chan, “Physically-attractive males increase men’s financial risk-taking,” *Evolution and Human Behavior*, vol. 36, no. 5, pp. 407–413, 2015.
[6] T. Tanimura, E. Jaramillo, D. Weil, M. Raviglione, and K. Lönnroth, “Financial burden for tuberculosis patients in low-and middle-income countries: A systematic review,” *European Respiratory Journal*, vol. 43, no. 6, pp. 1763–1775, 2014.
[7] T. K. Scharding, “Imprudence and immorality: A Kantian approach to the ethics of financial risk,” *Business Ethics Quarterly*, vol. 25, no. 2, pp. 243–265, 2015.
[8] J. Su and E. Furman, “A form of multivariate Pareto distribution with applications to financial risk measurement,” *ASTIN Bulletin*, vol. 47, no. 1, pp. 331–357, 2017.
[9] G. W. Y. Wang, S. H. Woo, and J. Mileski, “The relative efficiency and financial risk assessment of shipping companies,” *Maritime Policy & Management*, vol. 41, no. 7, pp. 651–666, 2014.
[10] J. Linarelli, “Luck, justice and systemic financial risk,” *Journal of Applied Philosophy*, vol. 34, no. 3, pp. 331–352, 2017.
[11] P. Allen and C. Petsoulas, “Pricing in the English NHS quasi market: A national study of the allocation of financial risk through contracts,” *Public Money & Management*, vol. 36, no. 5, pp. 341–348, 2016.
[12] J. Thomä and H. Chenet, “Transition risks and market failure: A theoretical discourse on why financial models and

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9870482, 1 page
https://doi.org/10.1155/2023/9870482

Retraction

Retracted: Singular Signal Measurement Based on Bidirectional Recursive Complex-Valued Wavelet Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Lv, J. Xie, and J. Huang, "Singular Signal Measurement Based on Bidirectional Recursive Complex-Valued Wavelet Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 9013770, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9013770, 10 pages
https://doi.org/10.1155/2022/9013770

Research Article

Singular Signal Measurement Based on Bidirectional Recursive Complex-Valued Wavelet Algorithm

Weiping Lv, Jinshan Xie, and Jing Huang

College of Mathematics and Information Engineering, Longyan University, Longyan 364012, China

Correspondence should be addressed to Weiping Lv; 82008013@lyun.edu.cn

Received 26 February 2022; Revised 12 April 2022; Accepted 18 April 2022; Published 14 May 2022

Academic Editor: Chin-Ling Chen

The normal operation fault of the power system is usually caused by a short-circuit fault. At this time, the system changes drastically from one state to another, accompanied by complex transient phenomena. Therefore, the measured signal contains a large number of transient components. How to effectively analyze such signals, extract their characteristics, and develop new protection devices has always been an important research field in power system protection technology. The protection of the power system is to achieve the purpose of correct action and elimination of faults by quickly detecting and locating faults. At present, the power signal analysis tools used in microcomputer protection include FFT, Kalman filter, and finite impulse response filter. They are efficient for the analysis of stationary signals, but have their limitations in analyzing nonstationary signals; especially it is difficult to identify nonlinear faults, such as the detection of high-impedance nonlinear short-circuit faults, which is a long-term unsolved problem in power systems. Based on wavelet transform, this paper selects complex-valued wavelet algorithm, analyzes a real-time recursive wavelet algorithm, and deduces the realization process of the algorithm in detail. The algorithm greatly reduces the computational complexity of the existing two-way recursive algorithm, can be used for real-time detection of fault signals in various fields of power system, and can be extended to realize other fast recursive algorithms of wavelet functions. Based on the sensitivity of complex-valued wavelet transform phase information to singularity, a method for real-time monitoring of power system fault mutation signals using the phase information of complex-valued wavelet fast recursion algorithm to assist amplitude information is proposed. The validity and practicability of this complex-valued wavelet and its real-time recursive algorithm for fault detection are demonstrated by an example.

1. Introduction

A singularity signal refers to a signal in which the signal itself or a certain order of its derivative has a sudden change at a certain moment, and the singularity detection is to identify the singularity of the signal and determine the degree of singularity. For a long time, the Fourier transform has been the main tool for studying the singularity of functions (signals). However, due to the lack of spatial local characteristics of the Fourier transform, it can only determine the overall nature of the singularity of a function, and it is difficult to determine the location and location of the singularity in space. Distribution and the singularity or sudden change of the signal is the most critical and important property of nonstationary signals in many cases. For example, mechanical failures, power system failures, abnormalities in the electrocardiogram of the EEG, and so on, all correspond to the mutation points of the test signal. Although they occur in different backgrounds, if the measured data is viewed as a signal, they are all reflected in the how to extract the location of the mutation point in the signal and determine its singularity (or smoothness). Therefore, the detection of signal singularity has particularly important practical significance. Fourier transform is an ideal tool for studying stable signals. For a long time, Fourier transform is also the main tool to study the singularity of signals. Transformation lacks spatial locality. It can only determine the overall nature of a signal's singularity, but it is difficult to determine the location and distribution of singular points in space. In fact, the singular points and irregular sudden changes in the signal often carry more important signals, which are one of the important characteristics of the signal. Wavelet analysis is an emerging discipline that has developed rapidly in recent years. It breaks through the Fourier transform without any resolution limitation in the time domain and can analyze the signal components in a specified frequency band and time period. It has good localization properties in both the time domain and the frequency domain. Wavelet transform can accurately capture the characteristics of transient signals and uses gradually fine time domain or spatial sampling steps for frequency components so that any detail of the signal can be focused on. Vividly speaking, it is to analyze the signal in a "slice type" and observe the evolution and characteristics of the signal in different frequency bands [1–10]. For example, detecting high-impedance faults (low-current arc faults) is a long-term unsolved problem in the power system. Traditional overcurrent protection devices simply cannot be detected; and some faults include arcs, which cannot form a fixed ground, resulting in a very small current, and sometimes there is no current at all.

A notable feature of power system failures is the sharp increase in current, which damages electrical equipment in terms of force and heat. Fuse protection and overcurrent protection are the original protection principles that reflect the characteristics of the sharp increase in current. Another feature of the fault is the sharp drop in voltage and the corresponding low-voltage protection. A protection principle that reflects both the voltage drop and the current increase is impedance (distance) protection. It reflects the distance to the fault by the decrease in impedance. Decide on the protection action: in order to more accurately distinguish between normal operating conditions and fault (or abnormal) conditions, there are no or few fault conditions during normal operation but there are large electrical quantities, such as negative sequence or zero sequence current, voltage, and power judgment: the use of power system fault protection is not limited to electrical quantities, but also other physical quantities, such as a large amount of gas and an increase in oil flow rate or oil pressure caused by a fault in the transformer tank [11–15]. The fault diagnosis is shown in Figure 1. This will cause the fault to last for hours or even days. In many cases, it can only be found by observation. This type of fault signal is irregular in time. The fault level may be quite high in a few cycles, but it drops to a normal load level in many cycles.

However, they generally have problems in identifying nonstationary time-varying signals and abrupt signals caused by nonlinear faults, which are difficult or even misidentified. As a result, fires, electrical breakdowns, personal accidents, burnout of switchgear, and damage to other devices are caused [16–22]. Therefore, in order to ensure the safe and reliable operation of the power system, it is necessary to conduct state monitoring and fault diagnosis of the power equipment. According to the various signals measured during the operation of the power equipment, the operating status can be determined through signal analysis [23–25]. The moment of the sudden change of the equipment fault information and the magnitude of the sudden change are captured, which helps to take early measures to restore the system to normal at the beginning of the fault. For fault identification, it is often through the analysis of information such as the amplitude and frequency components of the fault signal, combined with the existing fault signs of the power equipment to make judgments, which is convenient for fault location and rapid maintenance. The phase information of wavelet transform is often more sensitive to singularity, easy to capture singular points, and accurately detect signal mutations. Therefore, complex-valued wavelets can be used to calculate wavelet coefficients, but each calculation needs to complete a complete integral. The amount of calculation is with the rapid increase in the amount of data. Therefore, seeking a fast algorithm that satisfies the real-time requirements and retains the original calculation accuracy is extremely important for quickly capturing abnormal information of equipment, issuing forecasts in the early stage of equipment failure, and improving equipment operation reliability.

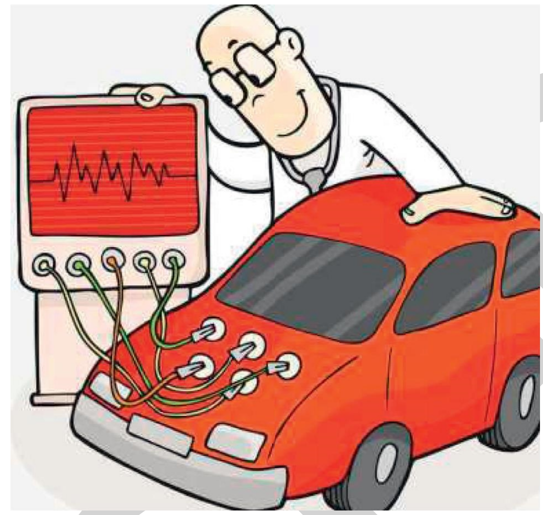


Figure 1: Fault diagnosis.

2. Real-Time Two-Way Recursive Complex Wavelet Algorithm

Singularities and irregular mutations in signals often contain important signals, which are one of the important characteristics of signals, such as spikes and spikes in eeg, edges of objects in images, rock ultrasonic detection signals with time-varying spectral characteristics and mechanical fault detection signals, and so on. The application of wavelet in power system can be expressed in the following aspects.

2.1. Status Monitoring and Fault Diagnosis of Power Equipment. Power equipment status monitoring and fault diagnosis is to decompose and process various electromagnetic, mechanical, and other physical signals generated by the basic equipment of the power system during operation and to determine its status in real time in order to issue an alarm at the initial stage of the fault or in the event of a fault. The electromagnetic signals emitted by power equipment are relatively stable during normal operation, and abnormal conditions in a day must contain singularities.

2.2. Harmonic Analysis of Power System. When a fault occurs in the power system, it is accompanied by the generation of high-order harmonics; in the high-voltage direct current transmission system, both the AC side and the DC side of the converter station produce high-order harmonics. In order to avoid the adverse effects of these harmonics, it is necessary to analyze and suppress them. Wavelet analysis transforms and projects this type of signal to different scales, which will clearly show the characteristics of these high-frequency and singular high-order harmonic signals. In particular, the wavelet packet has the characteristics of further subdividing the frequency space, which will be very good. Suppress high-order harmonics and provide a reliable basis.

2.3. Power System Transient Stability. When the power system is subject to large disturbances, various electromagnetic signal parameters that characterize the operating state of the system will undergo rapid changes and oscillations. To analyze such a sudden and localized signal, wavelet analysis is undoubtedly a good choice. The ability of wavelet analysis to capture and process weak mutation signals can be used in power system transient stability prediction research based on weak signals. Its "local refinement and amplification" feature can identify and track weak mutations in system variables, and then infer that mutations are caused. The time and location of local failures can improve the real-time and accuracy of power system transient stability prediction.

2.4. Dynamic Safety Analysis of Power System. When the power system is disturbed, it will cause the system voltage to fluctuate. Too low voltage will endanger the stability of the power system and may cause "voltage avalanche" in severe cases. Therefore, it is increasingly important to study the dynamic response of voltage. When the system is disturbed, it generates a voltage mutation signal. Apply wavelet analysis to decompose this abrupt signal into different scales, and then analyze the magnitude and phase of the abrupt signal on these different scales to determine the dynamic and safe operation of the power system.

2.5. Antielectromagnetic Interference. The large number of electromagnetic interference signals generated by the power grid makes it difficult to extract the characteristic signals of the operation behavior of power equipment. Wavelet analysis can apply wavelet transform to decompose the mixed signal containing the desired signal and electromagnetic interference signal to different scales. The wavelet coefficients associated with the signal are set to zero (to clear the interference signal), and then the reconstruction formula is used to construct the required signal, which also realizes the separation of the required signal and the interference signal and achieves the purpose of antielectromagnetic interference. The use of wavelet analysis to filter out the white noise in the signal has been successfully applied.

2.6. Fault Location of Transmission Line. Reliable operation of the power system requires timely and accurate knowledge of the fault location. Existing fault location methods and fault locators have been able to achieve this function, but there are still some problems in the processing of fault signals. If the current and voltage signals are obtained through fault recording, wavelet transform is used to decompose such singular and instantaneous fault signals, and the fault signals are clearly reflected on different scales, thus constructing the distance function (distance function). function), and then infer the time and location of the fault that caused the sudden change signal, and finally reflect the fault distance to achieve the purpose of fault location, which will improve the accuracy of fault location.

2.7. Short-Term Load Forecasting of Power System. Short-term load forecasting of the power system is a prerequisite for the economical and reliable operation of the power system. The various existing forecasting methods mainly include time series neural network method. The electric load has a special periodicity. The load fluctuates in cycles of days, weeks, and years, and small cycles are nested in large cycles. Wavelet transform can decompose various interleaved mixed signals composed of different frequencies into block signals on different frequency bands. Therefore, the wavelet transform of the load sequence can project the load sequence on different scales, and each scale can be approximated; the ground is regarded as different "frequency bands" so that the subsequences on each scale represent the components of different "frequency domains" in the original sequence, and they more clearly show the periodicity of the load sequence.

2.8. HVDC Transmission System. The analysis and detection of the bridge arm short circuit and commutation failure of the HVDC transmission system can still be realized by means of wavelet analysis, which is sensitive to singular, weak, and transient fault signals. In addition, wavelet analysis has equally important applications in HVDC system fault diagnosis, main equipment status monitoring, HVDC transmission line fault location, antielectromagnetic interference, and many other aspects.

Generally speaking, the expression of complex wavelet is

$$\psi(t) = \left|1 + \sigma|t| + \frac{\sigma^2}{2}t^2\right| e^{-\sigma|t|} e^{jw_0 t}, \tag{1}$$

where $w_0 = 2\pi$, $\sigma = 2\pi/3$. According to this, a two-way recursive algorithm (recursive wavelet transform) can be constructed. For a certain scale s, only 5 complex coefficients such as δ_1 and λ_1 and their conjugates need to be calculated once, and then only the wavelet coefficients need to be calculated recursively. With the small increase in data, each calculation of a wavelet coefficient requires 36 real number multiplications and 35 real number additions. Compared with direct integral transformation, the amount of calculation is greatly reduced, so it can be used for fast calculation. But it is composed of two parts, causal and noncausal, which need to calculate backward and forward recursively, respectively. The result of wavelet transform is not easy to be stable, and the calculation speed is reduced. Therefore, this paper has made further improvements on this basis and established a lower-order wavelet function, which only needs to perform backward one-way recursive calculation, which overcomes the deficiencies of the above methods. The repeated and irregular occurrence of these sudden fault currents results in a very low average RS current level. Even if the general protection device is set low enough, it will be affected by the abovementioned irregular characteristics and refuse to operate. However, a large amount of electric energy flows continuously to the point of failure. There are functions:

$$\psi_1(t) = \left|-\frac{6}{5}\sigma t - \frac{6}{5}\sigma^2 t^2 - \sigma^3 t^3\right| e^{\sigma t} e^{jw_0 t} u(-t). \tag{2}$$

Then, the wavelet function can be obtained:

$$\psi(t) = \left|\frac{6}{5}\sigma t - \frac{6}{5}\sigma^2 t^2 + \sigma^3 t^3\right| e^{-\sigma t} e^{jw_0 t} u(t), \tag{3}$$

where it satisfies

$$\psi(t) = \overline{\psi}_1(-t). \tag{4}$$

Its Fourier transform is

$$\widehat{\psi}(w) = \frac{6}{5}\cdot\frac{4\sigma^3 - \sigma(w - w_0)^2}{[\sigma + j(w - w_0)]^4}, \tag{5}$$

where $w_0 = 2\pi$, $f = 1/s$, $\sigma = \pi$. The wavelet function $\psi(t)$ satisfies the admissibility condition, and its time-domain and frequency-domain waveforms are shown in Figure 2.

$$\widehat{\psi}(0) = 0. \tag{6}$$

For the selected wavelet $\psi(t)$, the integral wavelet transform of the continuous signal $f(t)$ is

$$W_f(s, x) = \frac{1}{s}\int_{-\infty}^{\infty} f(t)\psi\left|\frac{x - t}{s}\right| \mathrm{d}t. \tag{7}$$

From the above formula, the discrete wavelet transform of the digital signal is shown in the following formula:

$$W_f(s, kT_s) = \frac{T_s}{s}\left| f(kT)\psi\left|\frac{kT_s}{s}\right|\right|, \tag{8}$$

where $f(nT_s)$ is the digital signal sequence (T_s is the sampling period) and ψ is the mother wavelet discrete sequence.

From the Z-transform of the sequence and its time-domain convolution properties, the Z-transform of the signal sequence, mother wavelet sequence, and wavelet coefficient sequence is

$$W(z) = \frac{T_s}{s}\left| f(z)\psi(z)\right|. \tag{9}$$

For wavelet sequence,

$$\psi\left|\frac{kT_s}{s}\right| = \left|\frac{6\sigma k T_s}{5s} - \frac{6}{5}\left|\frac{\sigma k T_s}{s}\right|^2 + \left|\frac{\sigma k T_s}{s}\right|^3\right| e^{-\sigma T_s k/s} e^{jw_0 T_s k/s}. \tag{10}$$

Let

$$\begin{cases} \beta = \dfrac{\sigma T_s}{s}, \\ A = -\left|\dfrac{\sigma T_s}{s} - j\dfrac{w_0 T_s}{s}\right|. \end{cases} \tag{11}$$

Finally, the recurrence formula obtained by the inverse Z-transformation is

$$\begin{aligned} W(s, k) = {} & \frac{T_s}{s}\{\delta_1 f((k-1)T_s) + \delta_2 f((k-2)T_2) + \delta_3 f((k-3)T_s)\} - \lambda_1 W(s, (k-1)T_s) \\ & - \lambda_2 W(s, (k-2)T_s) - \lambda_3 W(s, (k-3)T_s) - \lambda_4 W(s, (k-4)T_s). \end{aligned} \tag{12}$$

On the basis of this algorithm, only 4 wavelet transform initial values need to be calculated, and then all wavelet coefficients can be calculated by backward one-way recursion, which avoids forward recursive operations and reduces multiplication and addition operations by half. The required coefficients (δ, λ, etc.) only need to be calculated once and have a lower power calculation; this algorithm can meet the needs of real-time calculation. Use wavelet analysis theory to do multiresolution analysis (MRA) on the obtained singular electromagnetic signal, and decompose the signal into different scales.

Different wavelets show different characteristics in terms of orthogonality, tight support, smoothness, and even symmetry, but it is difficult to construct a wavelet function with four characteristics at the same time. Daubechies proved that when the scale function and wavelet function determined by a multiresolution analysis are real functions and both have compact support, the wavelet function no longer has symmetry or antisymmetric (except for Haar wavelet). In practical applications, only according to the needs of different signal decomposition, a compromise is made between several characteristics, and the wavelet that meets the needs is selected for decomposition. Qualitatively speaking, when the oscillation frequency of the detected signal is close to the oscillation frequency of the wavelet function of the corresponding scale, the signal obtains a wavelet decomposition with larger coefficients. This is the reason why wavelet analysis can extract different frequency

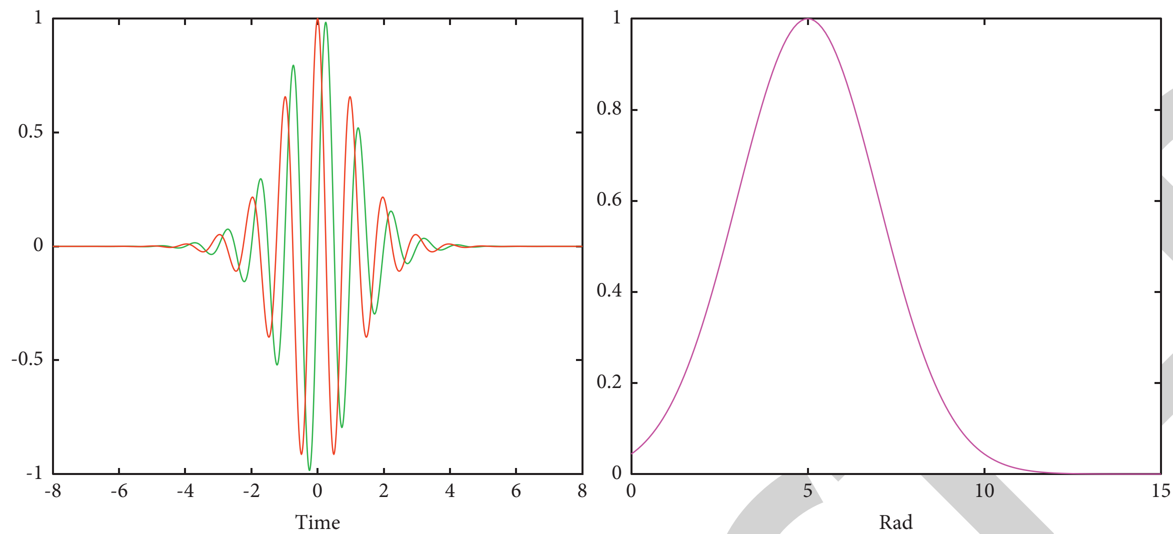


Figure 2: Time domain, frequency domain waveform.

components of the signal at multiple scales. In other words, the "entropy" value is usually used to measure the distance between the signal and the wavelet basis. The smaller the distance (i.e., the smaller the entropy value), the smaller the difference between the signal and the basis, and the signal gets the maximum decomposition. Therefore, different wavelets need to be selected for different transient signals. By comparing the entropy value of different wavelet decompositions, the wavelet with smaller entropy value is selected to obtain larger decomposition and achieve better detection effect.

The idea of the above recursive algorithm can also be extended to other wavelets with the same form. As long as the selected wavelet satisfies the form of the product of polynomial and exponential decay, its Z-transform can be transformed into a rational function of z-1, and the corresponding structure is constructed according to the aforementioned method. For the real-time recursive algorithm, the difference lies in the calculation of the coefficients. The predicted data is shown in Figure 3. The components on each scale reflect the different frequency components of the original signal, which can show the fault signal to reach the state the purpose of monitoring or fault diagnosis. This method has been successfully applied in the diagnosis of broken rotor bars in electric motors.

3. Real-Time Two-Way Recursive Complex Wavelet Algorithm in Power System Singular Signal Measurement

EMTP (Electromagnetic Transients Program) is an electromagnetic transient analysis program. It was originally created by Professor HW Dommel of the University of British Columbia (UBC) in Canada, and then it has been improved by the joint efforts of many experts. The mathematical model of EMTP is based on the phase coordinate system. It has a basic idea, which is to treat the three-phase components widely existing in the power system as three single-phase components with electromagnetic coupling. Such a processing method makes the simulation program have great flexibility. The program is mainly used for computer simulation of the electromagnetic transient process of the power system. The program can solve linear and nonlinear resistance, inductance, and capacitance circuits including lumped parameters, multiphase π-type circuits, multiphase distributed parameter circuits, various types of switches, the steady-state or transient process of a large-scale power system composed of transformers, power supplies, and control systems. In addition, all other systems that can be simulated by circuits, such as the steady-state or transient processes of mechanical systems, can be calculated with this program. The EMTP program has the advantages of large scale, strong function, and real simulation. It has been widely used in my country's institutions of higher learning, scientific research, design, and manufacturing and has been used in the research of some national key projects. On this basis, different subload sequences are respectively predicted. Since the periodicity of each subsequence is more obvious, it is obvious that the prediction results of the periodic autoregressive model (PAR) are more accurate.

Now the power system electromagnetic transient program (EMTP) based on offline calculation is becoming more and more powerful. The scale of the computing system network is getting larger and larger, and it can simulate the power system with thousands of nodes; there are many components that can be simulated, such as various power electronic devices such as SVC, SVG, TCSC, and AC/DC converter station; it can also simulate the internal fault of the transformer, the inrush current of the transformer, and the transient process of the current transformer (TA) and saturation voltage transformer (TV). Compared with the all-digital simulator based on real-time calculation, offline EMTP can not only simulate all the situations that the all-digital real-time simulator can simulate, but also can simulate some situations that it cannot simulate, and has some all-digital real-time simulators that do not have. Advantages include low price and low hardware requirements.

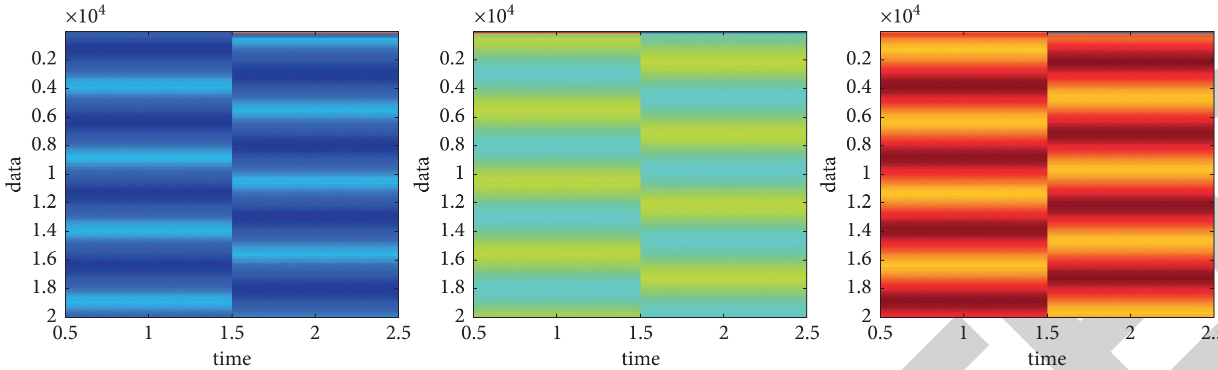


Figure 3: Predicted data.

Therefore, it is very meaningful if you can make full use of the powerful functions of offline EMTF to test relay protection. Finally, through sequence reconstruction, a complete hourly load forecast result is obtained, and its accuracy is greatly improved compared to directly using the original load sequence to predict.

Figure 4 (left one) shows the measured *a*-phase electric music waveforms when the *a*-phase grounding short circuit via linear resistance and the nonlinear resistance grounding short-circuit occur at the midpoint of the system transmission line. Figure 4 (right one) shows the three-phase voltage waveforms measured when a three-phase voltage short-circuited by linear resistance and nonlinear resistance occurs at the midpoint of the system transmission line. It also shows the measured voltage waveforms of the two phases *ab* when the two-phase *ab* is short-circuited to the ground via a linear resistance and the two-phase *ab* is short-circuited to the ground via a nonlinear resistance at the midpoint of the transmission line of the system. Besides, it also shows the measured three-phase voltage waveforms when a three-phase short-circuit to the ground via a nonlinear resistance occurs at the midpoint of the transmission line of the system.

Traditionally, Fourier analysis is used for dynamic and transient signals in power systems. But with the deepening of research, the Fourier transform can no longer meet the requirements of the sudden change of the signal. The emerging wavelet transform provides a powerful tool for sudden singularity signals and has been successfully applied to many fields. In the fault diagnosis, the identification of the fault is often through the analysis of the fault signal amplitude and frequency components and other information, combined with the existing fault signs of the fault system to make judgments, which is convenient for fault location and rapid maintenance. The phase information of wavelet transform is often more sensitive to singular signals, easy to capture singular points, and accurately detect signal mutations, so complex wavelets can be used to calculate wavelet coefficients. Complex wavelet transform can not only provide amplitude information, identify feature scale and its position on the time axis, but also provide phase information, which can be used to locate the singular point of the process. Therefore, the use of complex wavelet will help with the fault diagnosis rate. For the measured three-phase voltage signal, this paper uses wavelet transform to extract its characteristics.

The so-called feature extraction is to find indicators that can effectively reflect the signal so as to further identify and classify the signal. The wavelet analysis method has a strong feature extraction function, which can extract the fault features of the fault phase. As long as the various current and voltage signals are collected at a suitable sampling frequency, and each sampled value is subjected to wavelet transformation, it can be found that at the moment of the fault, the wavelet transformation of the fault phase current and the fault phase voltage has a modulus maximum value. Apparently, the characteristics of the fault are obvious; the wavelet transform of the nonfault phase current and the nonfault phase voltage has no sudden modulus maximum. Therefore, the wavelet transform of current or voltage can be used to constitute a phase selection element reflecting the transient component of the fault because the modulus maximum value of the wavelet transform only appears at the moment of the fault. The wavelet component obtained by the dyadic wavelet transform of the signal cannot be used as the characteristic quantity because although it contains all the information of the signal, the amount of data is too large to facilitate the online identification of the signal, and the edge of the signal wavelet transform often contains the most important part of the signal. Important features, and the amount of data is small, so it is more suitable for signal recognition. The algorithm is shown in Figure 5.

In the process of feature extraction, there is one point that needs special explanation. Both theory and actual decomposition prove that the position of the modulus maximum point of wavelet analysis is not constant on adjacent scales. However, if two modulus maximum points on two adjacent scales belong to the same modulus maximum line in the scale space (S, x), then it is considered that the modulus maximum points propagate along the scale. At the same time, the choice of wavelet base is also very important. The evaluation is shown in Figure 6.

To illustrate the effectiveness of the above real-time algorithm for detecting signal mutations, the following takes the voltage amplitude mutation shown in Figure 4 and the voltage harmonic distortion signal shown in Figure 7 as examples for analysis ($fs = 2000$ Hz). The wavelet and one-way recursive algorithms of this paper are used to decompose the wavelet, respectively, and the amplitude and phase information of the

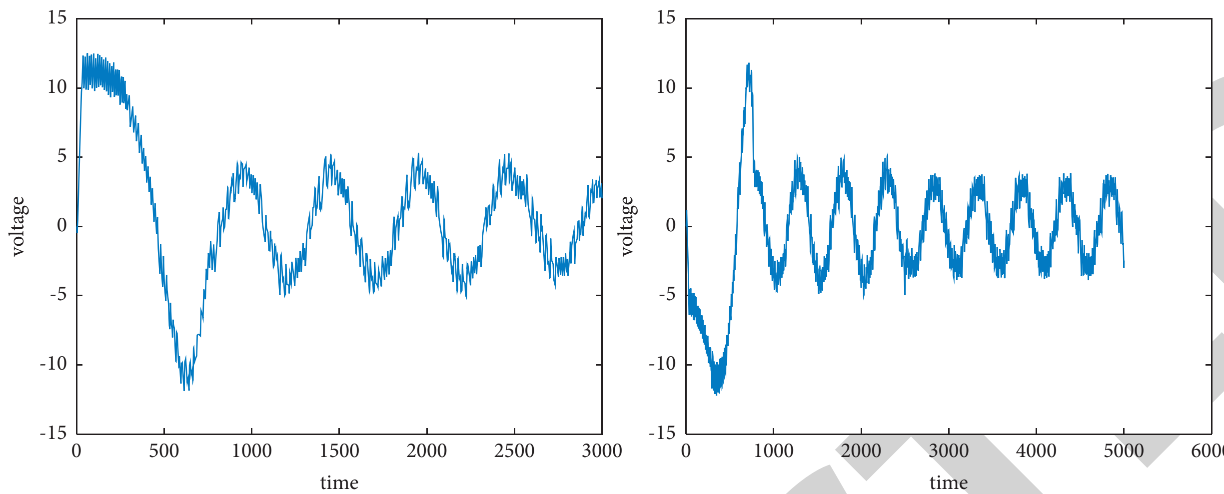


Figure 4: Measured a-phase electric music waveforms when the a-phase grounding short-circuit.

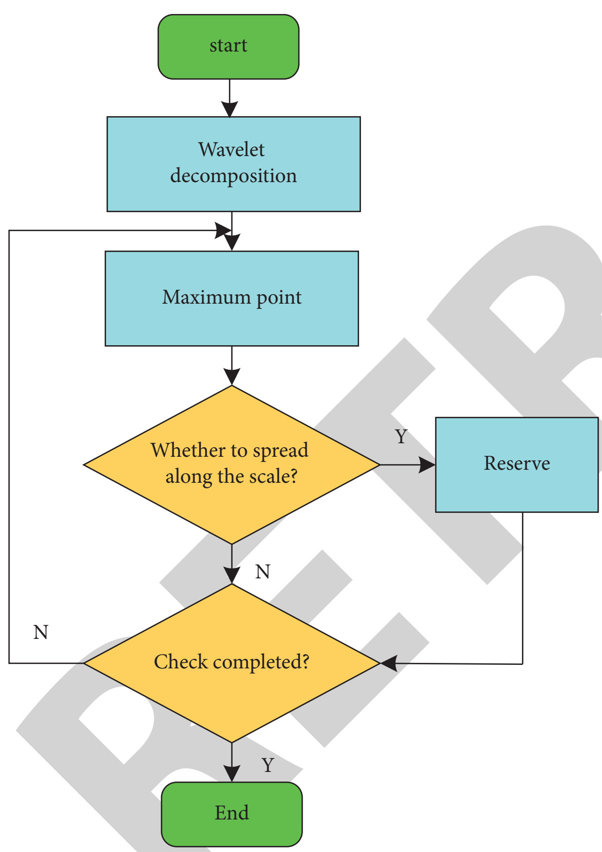


Figure 5: The algorithm.

wavelet transform are extracted, and the percentage errors of the calculation results of the two methods are compared. The abrupt amplitude signal shown in Figure 7 decreased by 2% in amplitude at 55 points, recovered in amplitude at 100 points, decreased by 1% at 150 points, and increased by 1% at 200 points, accompanied by 1% frequency deviation. Due to the small amplitude change, several signal amplitude change points cannot be detected from the original signal and its Fourier transform. Using the method in this paper and the literature to extract the amplitude and phase information of the wavelet transform, only two high-frequency scales with obvious effects $a = 0$ are listed in the figure. It can be clearly seen from the figure that the one-way recursive algorithm in this paper can still accurately locate the singularity of the signal from the two aspects of the wavelet transform amplitude and phase information. Compared with the wavelet transform in the previous literature, the suberrors can be controlled in the range of 5% or even lower, to achieve similar detection accuracy, and reduce the multiplication and addition operations by half compared with the method in the previous literature, and the decomposition result quickly reaches a stable value, so this recursive algorithm can be used to detect singularities in real time. As can be seen, the wavelet transform is very effective in signal localization and singularity analysis.

The harmonic distortion signal is that the power frequency signal contains high-order harmonics between 50 and 100 points, the harmonic distortion is 1.41%, and there are slight disturbances at 80 and 150 points. Due to the small harmonic content and disturbance, the sudden change information cannot be detected from the time domain signal and the Fourier spectrum. Also use the aforementioned two methods to perform wavelet transform on it, and extract the amplitude and phase information. It can also be clearly seen from the figure that the recursive algorithm in this paper can also accurately detect the singularity of the harmonic distortion signal from the wavelet transform amplitude and phase information and can also distinguish the disturbance in the harmonic component. At 80 points, the percentage error of the method in the previous literature can still be controlled in a low range, which once again reflects that the recursive algorithm in this paper not only greatly reduces the amount of calculation, but the accuracy of detecting distortion can also reach the previous literature. The predicted results are shown in Figure 8.

It should be noted that the wavelet and one-way recursive algorithm in this article can be widely used in other fields in addition to the above examples of detecting singularities. The author applies it to extract the wavelet ridgeline of the stator current when the rotor bar of the asynchronous motor is broken to distinguish the fault and obtains good results.

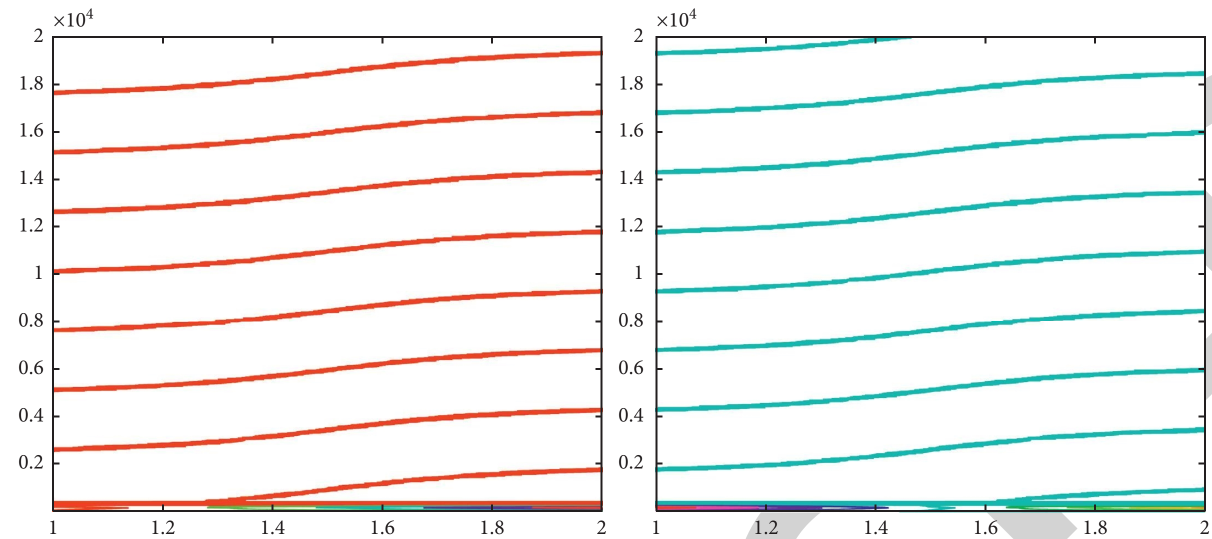


FIGURE 6: Evaluation.

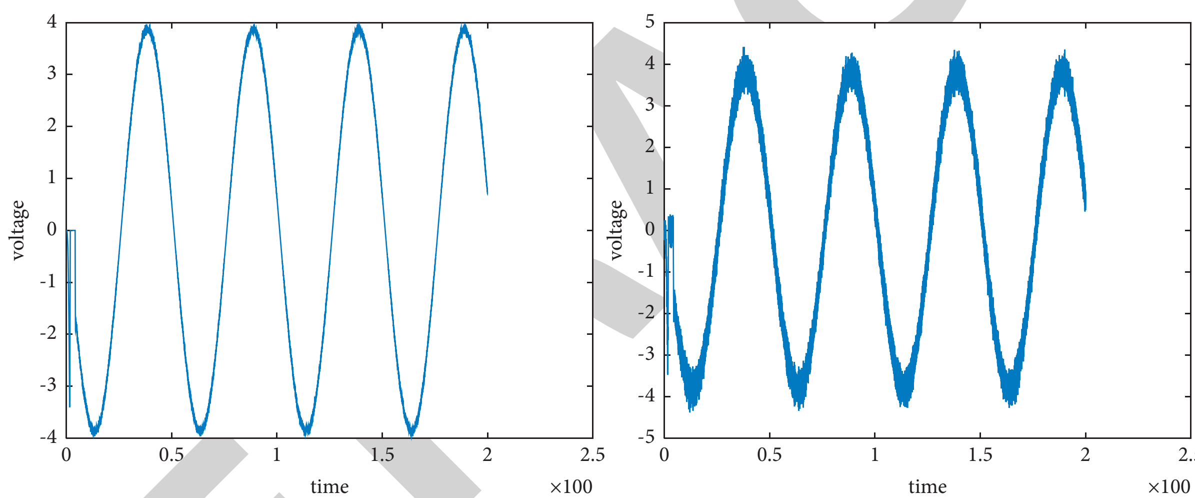


FIGURE 7: Voltage harmonic distortion signal.

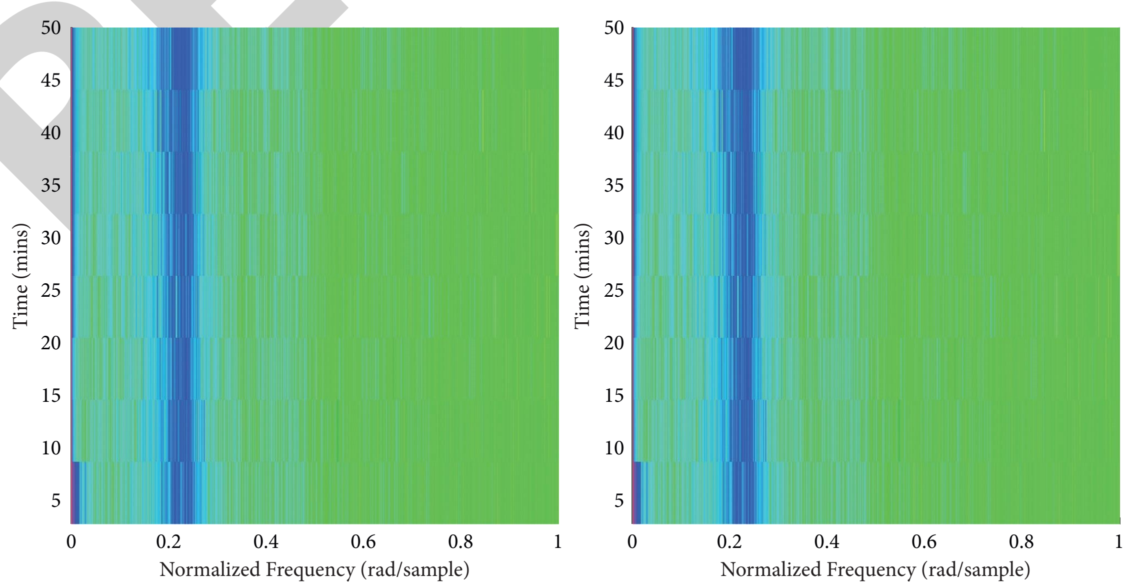


FIGURE 8: Predicted results.

4. Conclusion

Based on the integral wavelet transform, this paper analyzes a real-time recursive wavelet algorithm based on the selected complex wavelet as an example and derives the implementation process of the algorithm in detail. The algorithm greatly reduces the amount of calculation of the existing two-way recursive algorithm; thus, it can be used in real-time detection of fault signals in various fields of power system and can be extended to realize other fast recursive algorithms of wavelet functions. Based on the sensitivity of complex wavelet transform phase information to singularity, a method for real-time monitoring of power system fault mutation signals using phase information auxiliary amplitude information of complex wavelet fast recursive algorithm is proposed. An example is used to demonstrate the effectiveness and practicality of this complex wavelet and its real-time recursive algorithm to detect faults.

The work of this paper still needs to be further deepened. On the one hand, the selected complex wavelet has a two-way recursive algorithm. Although it can be calculated quickly, the algorithm includes two parts of forward and reverse recursive calculations. The reverse recursive part needs to wait until the data of the entire analysis window is completely obtained.It actually can be carried out, which is not conducive to the real-time requirements of fault diagnosis. In order to remove the backward recursive calculation and improve the real-time performance of diagnosis, it can be considered that only the positive time function is used to form the mother wavelet function.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Youth Project of the National Natural Science Foundation of China, Graph and Hypergraph Problem Based on Inequality Theory and Its Application (no. 11601214).

References

[1] L. Ge, Y. Li, J. Yan, Y. Wang, and N. Zhang, "Short-term load prediction of integrated energy system with wavelet neural network model based on improved particle swarm optimization and chaos optimization algorithm," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 6, pp. 1490–1499, 2021.

[2] R. Huang and J. Tian, "Wavelet-based elman neural network with the modified differential evolution algorithm for forecasting foreign exchange rates," *Journal of Systems Science & Information*, vol. 9, no. 4, pp. 421–439, 2021.

[3] I. B. Slimen, L. Boubchir, Z. Mbarki, L. Boubchir, Z. Mbarki, and H. Seddik, "EEG epileptic seizure detection and classification based on dual-tree complex wavelet transform and machine learning algorithms," *The Journal of Biomedical Research*, vol. 34, no. 3, pp. 151–161, 2020.

[4] Z. Khan and S. Amin, "Bottleneck model with heterogeneous information," *Transportation Research Part B: Methodological*, vol. 112, no. 1, pp. 157–190, 2018.

[5] E. Emad, A. Safey, A. Refaat, Z. Osama, E. Sayed, and E. Mohamed, "A secure image steganography algorithm based on least significant bit and integer wavelet transform," *Journal of Systems Engineering and Electronics*, vol. 29, no. 03, pp. 639–649, 2018.

[6] B. Shi, X. Yang, and L. Yan, "Optimization of a crude distillation unit using a combination of wavelet neural network and line-up competition algorithm," *Chinese Journal of Chemical Engineering*, vol. 25, no. 8, pp. 1013–1021, 2017.

[7] C. Liu and Y. S. B. S. Gong, "Bridge damage detection using spatiotemporal patterns extracted from dense sensor network," *Measurement Science and Technology*, vol. 28, no. 1, Article ID 014011, 2017.

[8] K. Ma and M. A. J. Z. Soltani, "Wind farm power optimization and fault ride-through under inter-turn short-circuit fault," *Energies*, vol. 14, no. 11, p. 3072, 2021.

[9] J. Zhang and H. Y. J. Wu, "Turn-off modes of silicon carbide MOSFETs for short-circuit fault protection," *Journal of Power Electronics*, vol. 21, no. 2, pp. 475–482, 2020.

[10] H. Shao, W. H. K. Lam, and M. L. Tam, "A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand," *Networks and Spatial Economics*, vol. 6, no. 3-4, pp. 173–204, 2006.

[11] C. Shi and X. Lu, "Online detection method for inter-turn short-circuit fault of permanent magnet synchronous motor based on deep learning," *IOP Conference Series: Earth and Environmental Science*, vol. 546, no. 5, Article ID 052015, 2020.

[12] A. L. Rodriguez, "Power transformer under short-circuit fault conditions: a multiphysics approach," *Transform Magazine*, vol. 7, no. 3, pp. 120–127, 2020.

[13] S.-M. Hosseininasab and S.-N. Shetab-Boushehri, "Integration of selecting and scheduling urban road construction projects as a time-dependent discrete network design problem," *European Journal of Operational Research*, vol. 246, no. 3, pp. 762–771, 2015.

[14] W. Shao, R. Wu, L. Ran, H. Jiang, and P. A. Mawby, "A power module for grid inverter with in-built short-circuit fault current capability," *IEEE Transactions on Power Electronics*, vol. 1, p. 1, 2020.

[15] X. Wang and P. Y. F. Wang, "Online estimation of short-circuit fault level in active distribution network," *Applied Sciences*, vol. 10, no. 11, pp. 3812–3819, 2020.

[16] K. A. Mohamed, B. Azeddine, and C. B. Djamal Eddine, "Short-circuit fault diagnosis of the DC-Link capacitor and its impact on an electrical drive system," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 2807–2814, 2020.

[17] H. Jia, J. Yin, T. Wei, Q. Huo, J. Li, and L. Wu, "Short-circuit fault current calculation method for the multi-terminal DC grid considering the DC circuit breaker," *Energies*, vol. 13, no. 6, pp. 232–243, 2020.

[18] J. Ouyang and M. D. Y. Y. Pang, "Improved voltage control method of power system based on doubly fed wind farm considering power coupling under grid short-circuit fault,"

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9851686, 1 page
https://doi.org/10.1155/2023/9851686

Retraction

Retracted: Machine Learning-Based Emotion Factor Analysis of Sport Fan Community

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Guo, B. Liu, and Z. Yang, "Machine Learning-Based Emotion Factor Analysis of Sport Fan Community," *Security and Communication Networks*, vol. 2022, Article ID 2674987, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2674987, 9 pages
https://doi.org/10.1155/2022/2674987

Research Article

Machine Learning-Based Emotion Factor Analysis of Sport Fan Community

Hengchuan Guo, Baoluo Liu , and Zhen Yang

Luoyang Institute of Science and Technology, Luoyang 471023, Henan, China

Correspondence should be addressed to Baoluo Liu; 200900500356@lit.edu.cn

Received 30 March 2022; Revised 25 April 2022; Accepted 29 April 2022; Published 14 May 2022

Academic Editor: Chin-Ling Chen

Football is one of the most popular sports in the world. As the popularity of football continues to grow worldwide, so does the number of incidents of violence on the pitch. Today, doping, match fixing, black whistles, and football hooliganism are ranked as the four most toxic aspects of sport. How to study the factors that cause aggressive behaviour of fans from a psychological perspective has become a key issue in the field of sports. Therefore, this study proposes a method for mining the psychological factors of sport fan community members based on machine learning clustering. Firstly, three different members of a large fan community, i.e., university students, office workers, and unemployed people, are used as research subjects to investigate the psychological factors influencing fans' aggressive behaviour using a questionnaire method. Secondly, the data obtained were mined and analysed using the K-means clustering algorithm in machine learning techniques. At the same time, a K-means initial clustering centre optimization algorithm based on principal component analysis (PCA) was proposed for the data characteristics of the interaction of psychological factors. The results show that the new algorithm significantly improves the quality of clustering compared with other optimization algorithms and accurately identifies the multiple factors that contribute to the occurrence of fan attacks.

1. Introduction

Football is the world's most popular, most widely played, and most influential sport and is known as the "world's number one sport." An exciting game of football attracts thousands of spectators and hundreds of millions of television viewers. The news about football takes up a lot of space in the world's newspapers and magazines. Today, football has become an integral part of people's lives. According to incomplete statistics, there are now about 800,000 teams playing regularly in the world and about 40 million registered players, including about 100,000 professional athletes [1–6].

The reason why football attracts such a large number of spectators and participants is ultimately down to the sport's appeal. People who play football regularly are conducive to developing good mental qualities, but from time to time we hear about an element of discord in football: aggressive behaviour by fans. When a fan's favourite team loses a match at the time, it can happen that the fans cannot accept the reality of being eliminated. Led by a few fans, the majority of fans follow suit, abusing the players, then attacking the referee, attacking the police, and burning police cars.

Faced with the constant and gradually escalating fan unrest, governments have tried to do everything possible [7–9]. For example, Spain banned Manchester United fans who had been drinking during the Champions Cup match between Manchester United and Real Madrid. Britain punished football hooligans for ten years for not being allowed to watch football. Germany blacklisted some repeat offenders and banned them from attending matches abroad. Belgium kept fans from all countries tightly segregated while conducting massive anti-riot drills. However, these methods have had little effect. What was it about the match that got the crowd so excited? The crowd for a high-level match can range from a few thousand to 50,000 or 60,000. The spectators vary in composition, from students to social workers, and some know a little bit about football and some know nothing about it, so they each go to the game with a different mentality [10–12]. What exactly is it that causes fans to repeatedly cause trouble during and after matches? Numerous scholars and experts at both home and abroad have made studies to address these aspects, but have never got to the root of such problems.

The main problem to be solved is to analyse the characteristics of fans' aggressive behaviour from the psychological point of view and try to use advanced machine learning technology to dig out what factors induce fans' aggressive behaviour from a large number of psychological questionnaire data, to carry out effective prevention. Clustering is an important method in machine learning [13–15] and an important element in data mining, which is widely used in many fields, including intelligent business, image pattern recognition, statistics and pattern analysis, and information retrieval. The K-means clustering algorithm [16] is widely used in various fields for its simplicity, ease of implementation, efficiency, and scalability. It is widely used in various fields.

Therefore, this study uses the K-means clustering algorithm to mine the required knowledge from a large amount of psychological questionnaire assessment data. Firstly, three different members of a large fan community, i.e., university students, office workers, and unemployed people, are used as research subjects, and a questionnaire is used to investigate the psychological factors affecting the aggressive behaviour of fans. Secondly, a modified K-means clustering algorithm from machine learning techniques was used to mine and analyse the obtained data. The experimental results validate the effectiveness of the proposed improved K-means clustering algorithm in mining the psychological factors of fan community members.

The rest of the study is organized as follows: in Section 2, a related research is studied in detail, while Section 3 provides the detailed improved K-means clustering algorithm based on PCA. Section 4 provides the detailed application of improved K-means clustering to psychological factors in fans. Section 5 provides the results and discussion. Finally, the study is concluded in Section 6.

2. Related Research

According to the mental venting perspective discussed by Freud and Lorenz, repressed aggressive forces may evolve into real aggression if they cannot be eliminated in some socially acceptable way. Bennett et al [17] analyse the problems created by football hooliganism from a sociological perspective. Romanet et al [18] point out the importance of the relationship between fans and the team they support. It is argued that stadium violence is actually created by those so-called fans. The so-called fans were not actually watching the game, but simply using the stadium as a place to vent their frustrations.

However, the conclusions drawn from the above studies are some personal empirical analyses and their reliability has yet to be verified. Furthermore, due to the limitations of manual processing, the sample size of the data analysed by these methods is small and cannot be applied to a large number of fan members. As a result, the conclusions drawn on psychological factors are somewhat controversial. As the basic tools of scientific research move from the traditional "theory + experiment" to the current "theory + experiment + computation," the importance of data mining and machine learning is becoming increasingly apparent. This is because the purpose of "computing" is often data analysis, and the core of data science is precisely the analysis of data to derive value. Clustering is an important element of data mining and is one of the fastest emerging areas of "new algorithms" in machine learning. Rochat et al. [19] used clustering algorithms to analyse the psychology of "swiping" on mobile apps. Kloos et al. [20] used a cluster randomization algorithm to deliver an online positive psychology intervention to nursing home staff. Aiyer et al. [21] proposed an analysis of psychological transition and neighbourhood relationships in adulthood using a multilevel clustering algorithm. Evidence was found for the direct effects of cluster membership and structural factors on neighbourhood relationships. It is evident from the above studies that clustering algorithms have greater potential for application in psychological factor analysis and mining.

The K-means algorithm, as a representative of division-based clustering algorithm, is again one of the top ten classical algorithms for data mining. In this study, the K-means clustering algorithm is used to mine the psychological factors of fan community members. Due to the complexity of fans' psychological factors, the initial clustering centre randomly selected in the typical K-means algorithm will lead to low clustering quality. The purpose of this study was to improve the K-means algorithm and use it to accurately dig out various factors that lead to fans' aggressive behaviour.

The main contributions of this study are as follows:

(1) A K-means initial cluster centre optimization algorithm based on principal components analysis (PCA) is proposed for the data characteristics of psychological factor interactions. Compared with other existing clustering algorithms, the new algorithm is more accurate in mining psychological factors.

(2) The improved K-means clustering algorithm was used to mine multiple causes of fan aggressive behaviour among members of the fan community and to obtain the proportion of each cause. Also, the occupational distribution of the sample of fan community members had to be mined. In addition, a summary analysis of the psychological factors that lead to the generation of fan violence is made.

3. Improved K-Means Clustering Algorithm Based on PCA

3.1. Basic Principles of K-Means Clustering. As a distance-based partition clustering algorithm, the K-means clustering algorithm has the advantages of simple algorithm structure, high running efficiency, and wide application range [22–24]. The K-means clustering algorithm is generally optimized by the objective function shown as follows:

$$E = \sum_{j=1}^{K} \sum_{x \in C_j} \left\| x - m_j \right\|^2. \quad (1)$$

It can be seen that the objective function shown in equation (1) is a sum-of-square error calculation process, where E is the clustering criterion function, K is the total number of clusters, C_j, $j = 1, 2, \ldots, K$ is the cluster j in the cluster, x is a clustering target in cluster C_j, and m_j is the average size of cluster C_j. The process of cluster analysis based on the K-means clustering algorithm is shown in Figure 1.

The input parameters of the K-means clustering algorithm are the value K and the number n of clustering targets in the dataset X. The output is the K clusters that minimize the clustering criterion function E. The basic flow of the K-means clustering algorithm is as follows [25]:

Step 1: Input the parameters and initialize the K clustering centres

Step 2: Calculate the value of E

Step 3: Update the centres of each cluster and calculate the new E

Step 4: Check whether the convergence condition is satisfied

Step 5: Output the parameters and finish if yes, and skip to Step 2 if no

All division-based clustering algorithm methods have a significant problem [26–28]: they are sensitive to the initial clustering centre. Therefore, for K-means clustering algorithms, the initial clustering centre is a key factor in determining the quality of clustering. A good initial clustering centre not only can effectively avoid the algorithm from falling into local optimum but also can greatly reduce the time overhead of the algorithm, thus improving the algorithm clustering quality and time efficiency. Due to the complexity and interaction of fan psychological factors, randomly selected initial clustering centres can lead to poor clustering quality.

3.2. Initial Clustering Centre Optimization Method. To optimize the initial clustering centres, it is first necessary to know what kind of centres are good initial clustering centres. Consider two extreme cases: (1) if the classification of the clusters is known in advance and the centre of each cluster is used as the initial clustering centre, then the algorithm only needs one iteration to obtain a very good clustering result; (2) if k sample points that are close together and at the edge of the entire sample space are chosen as the initial clustering centres, not only will many iterations be required, but the final clustering result is likely to fall into a local optimum.

Therefore, a good initial clustering centre should satisfy two conditions: (i) they are separated from each other by a certain distance and (ii) they are as close as possible to the centre of the intrinsically implied cluster. Existing distance-based and recursive-based refinements aim to make the initial clustering centres satisfy condition 1, while density-based refinements combine both condition 1 and condition 2. However, these methods introduce a number of additional parameters, which are often difficult to determine and are not borrowable when the datasets differ.

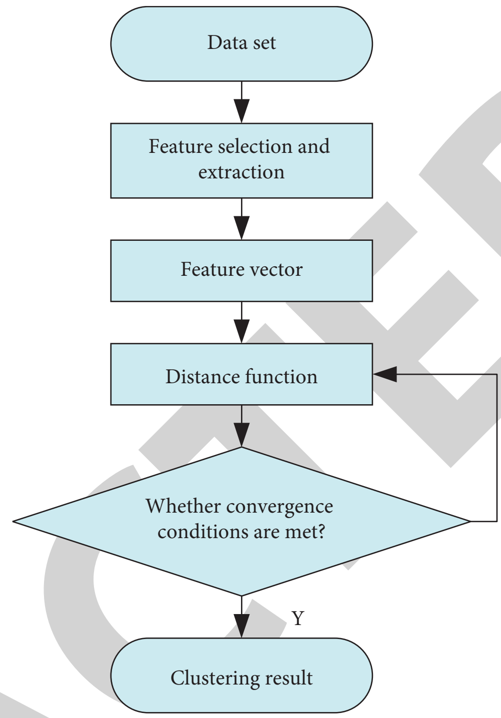


Figure 1: Cluster analysis schematic model.

How do you find the initial cluster centre so that it satisfies both conditions? Consider a simple average height problem: there are 30 students (each of whom knows his or her own height). How do you divide them into “tall” and “short” groups by height (assuming 15 students in each group) and find the approximate average height of each group? A quick and easy way to do this is to have 30 students sorted by height from shortest to tallest, and then, the 7th and 22nd students will be the approximate average height of each group.

Borrowing from the idea of solving the average height problem, for one-dimensional data, the initial cluster centres can be obtained by first sorting and then averaging the middle points. For high-dimensional data, you need to find a way to sort them and then take the same mean points to get the initial cluster centres. As the most commonly used linear dimension reduction method, principal component analysis (PCA) based on multivariate statistics has an excellent performance in feature extraction of high-dimensional data. The research of Iannucci [29] shows that PCA can reduce the dimension of the original features by projection without losing the information as much as possible.

Therefore, this study proposes a PCA-based K-means initial clustering centre optimization algorithm. The main idea of this algorithm is as follows: firstly, the high-dimensional data are reduced to one-dimensional data by principal component analysis; then the one-dimensional data are sorted in ascending order; then the one-dimensional data are clustered using the K-means algorithm; and finally, the initial clustering centre is obtained from the clustering results.

3.3. Steps in the Implementation of the Proposed Algorithm. The new algorithm uses the K-means clustering algorithm itself to divide the sorted data into k subsets, which is a good way to reduce the initial clustering centre bias caused by asymmetric clusters in the data sample (i.e., some clusters have more sample points and some clusters have fewer sample points).

The steps of the PCA-based initial clustering centre optimisation algorithm can be briefly described as follows:

Input: dataset D. Suppose it contains n data sample points, each containing a c-dimensional attribute, and the number of clusters to be divided is k.

Process:

Step 1: reduce the original multidimensional data to one-dimensional data using the PCA algorithm [30], denoted as $A_i\,(1 \le i \le n)$.

Step 2: sort the one-dimensional data A_i in ascending order.

Step 3: sort the sorted A_i into k clusters using the K-means clustering algorithm, where the k initial clustering centres are shown as follows [31]:

$$\text{centre}_j = A_i\left(i = \frac{2 \times j - 1}{2 \times k} \times n, \quad 1 \le j \le k\right). \qquad (2)$$

Step 4: divide the original data into k subsets based on the classified A_i and the one-to-one correspondence between A_i and the original multidimensional data.

Step 5: find the centroids of each of the k subsets.

Step 6: use the k sample points nearest to the centroids of the subsets in the original data as the initial clustering centres.

Output: k initial clustering centres.

3.4. Complexity Analysis. The main time spent in the principal component analysis is in finding the eigenvalues and eigenvectors. For an $n \times d$ matrix, it takes $O(nd)$ to calculate the covariance matrix and $O(d^3)$ to perform an eigenvalue decomposition on a $d \times d$ matrix. If the dataset is projected into the first m principal components, then only the first m eigenvalues and eigenvectors need to be found. This can be obtained using more efficient methods such as curtain calculation, which has a time complexity of $O(md^2)$, where n is the number of data samples, d is the data dimension, k is the number of clusters, and t is the number of generations selected.

The time complexity of the K-means clustering algorithm is $O(nkt)$. Since $m = 1$, the total time complexity of the PCA-based K-means initial clustering centre optimisation algorithm is $O(nd + d^2 + nkt)$.

4. Application of Improved K-Means Clustering Algorithm to Psychological Factors in Fans

4.1. Psychological Definition of Aggressive Behaviour. According to sport psychologists, aggression is a purposeful act of hurting another person, which can harm their physical or mental health through words or physical actions. In other words, aggression is an intentional action with the aim of causing harm or suffering. Sport psychologists believe that three factors influence the development of aggressive behaviour: firstly, the individual's innate tendency to have a reaction; secondly, the excitability factor in the mind; and thirdly, the experience factor. In short, we see aggression as an individual's reaction to the different levels of anger he or she experiences. A mental model of how aggression arises is shown in Figure 2.

4.2. Data Acquisition and Processing. Currently, the main methods of data collection for psychological aspects include expert interviews, literature, questionnaires, and mathematical statistics. In this study, three different members of a large fan community, namely university students, office workers, and unemployed people, were used as the research subjects, and the questionnaire method was used to study the psychological factors affecting the aggressive behaviour of fans. The total number of people in a large fan community was 2927, of which 383 were university students, 2017 were office workers, and 228 were unemployed. Questionnaires were distributed regarding the influencing factors that trigger fan disturbances to understand the causes and control methods of fan disturbances.

After data collection by the above method, fans need to be classified. Based on the purpose of watching football at the venue, fans can be classified into the following five categories: (1) knowledge-seeking type: the main motivation for this type of viewer is to know the outcome of the match; (2) aesthetic type: these are the fans who appreciate the game as a work of art and are as culturally sophisticated as the inquisitive fans; (3) entertainment type: this type of fan comes to the stadium to be entertained, to amuse themselves, and to spend their leisure time; (4) common-seeking type: there is a social psychology of seeking social affiliation and recognition from others; and (5) venturing frenzy type: this segment of fans is particularly fascinated by the heated atmosphere of the stadium. There is no single factor that causes aggressive behaviour in fans. A 12-factor questionnaire was developed for these five categories of fans, as shown in Table 1.

5. Experiment and Result Analysis

5.1. Manual Dataset Validation. To test the effectiveness of the improved PCA-based K-means clustering algorithm, a comparison with the typical K-means [29] and multilevel clustering [21] was performed on an artificially simulated dataset. The CPU used for the experiments was an Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz, with 4 GB of RAM, a 64 bit WIN 10 Operating System, and a MATLAB 2016b software environment. A total of three artificially simulated

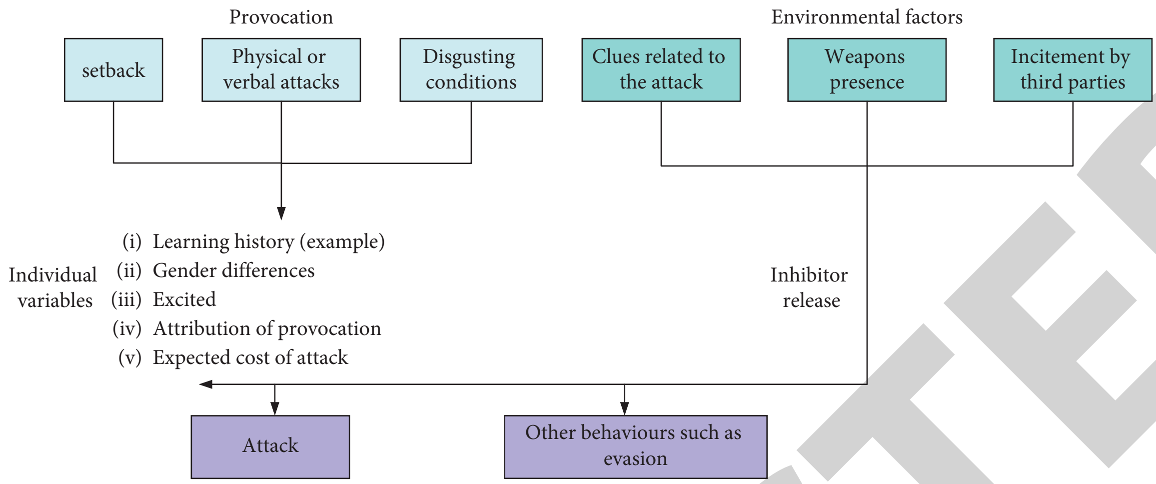


Figure 2: Psychological model for the production of aggressive behaviour.

Table 1: Questionnaire for the fan community (partial).

No.	Causes of aggressive behaviour by fans
1	The wrong cultural view of football
2	The sport of football itself and the characteristics of its playing process
3	Corruption in the racecourse
4	Moral attribution and player quality
5	Management factors
6	The negative role of the media
7	History of the relationship between the teams in the competition
8	Nature of the competition
9	Cultural bias and inadequate punishment
10	Modern social pressures
11	An increase in fan appreciation
12	Location of the competition
13	Others

datasets (randomly generated using a Gaussian normal distribution, $\sigma = 0.7$) were used, including simple and complex clustering structure features: Feature A, Feature B, and Feature C. The parameters of the artificially simulated datasets are shown in Table 2.

A two-dimensional scatter plot of the three datasets is shown in Figure 3. The experimental results of the improved K-means clustering algorithm on three artificial datasets are shown in Table 3.

As can be seen from Table 3, for both the manual dataset 1 and the manual dataset 2, the improved K-means algorithm obtained the smallest number of iterations and the sum of squared errors, indicating the highest quality of clustering. This indicates that the distance between the initial centre and the actual centre obtained by the improved K-means algorithm is the closest. For the artificial dataset 3, the multilevel clustering obtained the largest error sum of squares, the other 2 algorithms were the same, and thus the improved K-means algorithm had the highest quality of clustering. Although the improved K-means algorithm has more iterations than the multilevel clustering, the clustering quality of the multilevel clustering is poor.

Combining the three sets of manual simulation dataset experiments, it can be seen that the initial clustering centres obtained by the improved K-means algorithm are close to the actual centres of the data samples, and the clustering quality obtained is optimal in all cases, and the number of iterations is also the lowest. Thus, the effectiveness of the improved K-means algorithm is verified.

5.2. Clustering Mining Results. A total of 2000 questionnaires were distributed to different fans and 1815 questionnaires were returned, with a return rate of 90.8%. The data from the questionnaires were analysed by applying a modified K-means algorithm for cluster analysis. The dimension of cluster analysis is 12, which corresponds to the 12 factors of the questionnaire. Taking the factors with serial number 1, serial number 2, and serial number 3 in Table 1 as examples, the initial cluster centroids obtained from the modified K-means clustering are shown in Table 4. The final analysis results of the psychological factors about fans' aggressive behaviour after cluster analysis are shown in Table 5.

The greater the proportion of factors, the greater their influence on the aggressive behaviour of fans. Table 5 shows that of the 12 factors that influence the development of aggressive behaviour among fans, "modern social pressure" has the greatest weight, followed by "increased appreciation

Table 2: Parameters of the artificially simulated dataset.

Datasets	Structural features	Number of attributes	Sample size for the first category	Sample size for the second category	Sample size for the third category
1	A	2	60	60	60
2	B	2	100	50	50
3	C	2	50	25	50

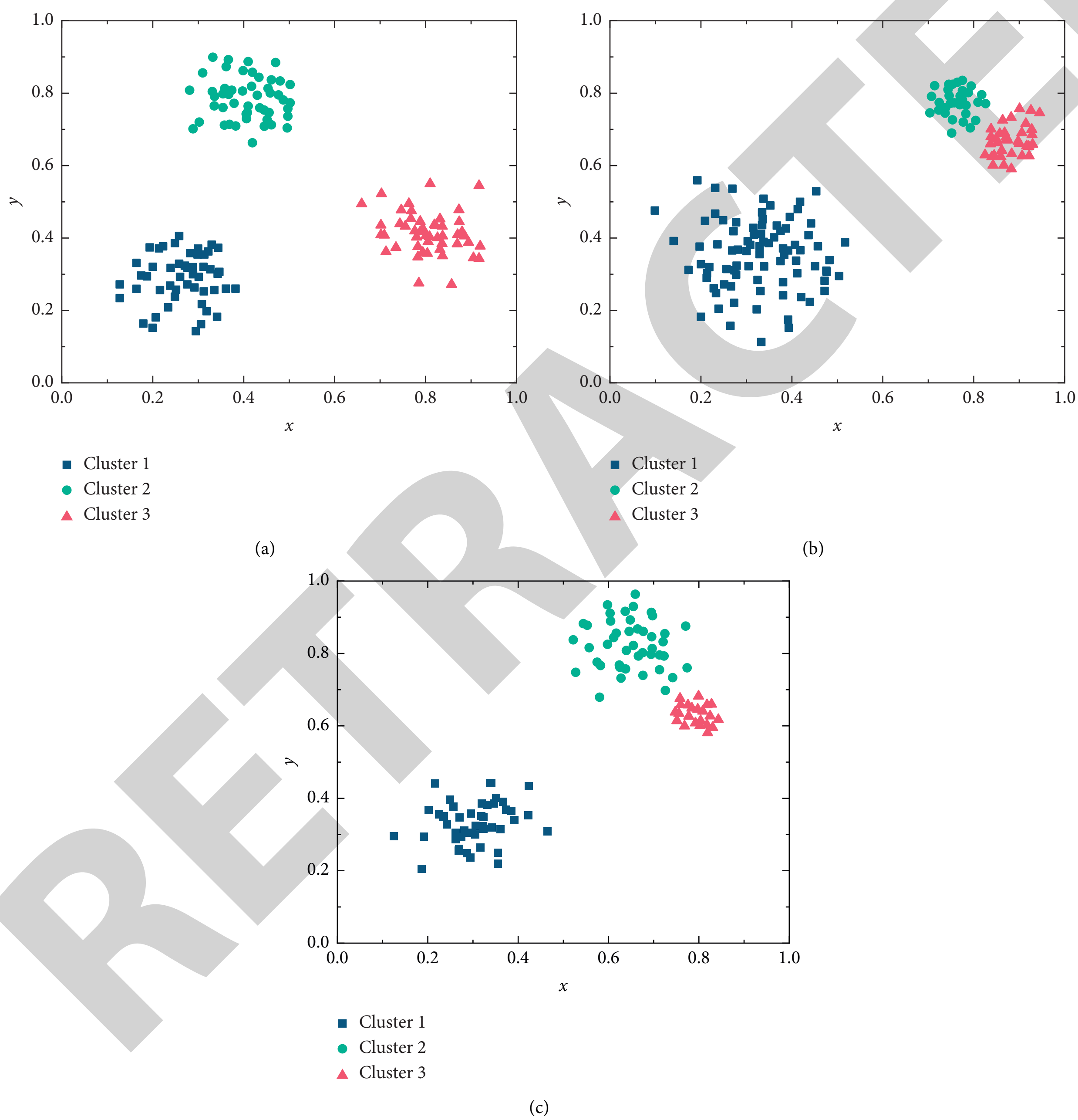


Figure 3: Two-dimensional plots of the three artificial datasets. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3.

Table 3: Comparison of clustering results on the manual dataset.

Datasets	Clustering algorithms	Number of iterations	Sum of squared errors within a cluster
1	Improving K-means	2	83.530
	K-means	2	86.056
	Multilevel clustering	2	84.647
2	Improving K-means	3	150.782
	K-means	5	167.653
	Multilevel clustering	3	159.031
3	Improving K-means	3	74.052
	K-means	5	74.052
	Multilevel clustering	2	109.581

Table 4: Initial cluster centroids obtained by improved K-means clustering.

No.	Initial cluster centroids
1	(1.463326, 0.5131198, 0.6760331, 0.1187017, 0.2534607, 0.6067924, 0.6952479, 0.1956907, 0.8017820, 0.5259556, 0.1564559, 0.42580766, 0.37190083, 0.3680441, 0.3744835, 0.09604106, 0.4188017, 0.3913223, 0.3090909, 0.6013688, 0.5996901, 0.6358471, 0.2923554)
2	(1.666667, 0.2166667, 0.4333333, 0.8736559, 0.2500000, 0.7500000, 0.6666667, 0.7823129, 0.4270833, 0.7604167, 0.8019324, 0.70454545, 0.09090909, 0.1444444, 0.4375000, 0.88172043, 0.4166667, 0.4250000, 0.6166667, 0.6562500, 0.6562500, 0.5729167, 0.8229167)
3	(2.000000, 0.0000000, 0.0000000, 0.0483871, 0.0000000, 0.0000000, 0.1875000, 0.0000000, 1.0000000, 0.6875000, 0.2173913, 0.04545455, 0.21212121, 0.2000000, 0.2500000, 0.00000000, 1.0000000, 1.0000000, 0.9000000, 0.9375000, 1.0000000, 1.0000000, 0.2812500)

Table 5: Analysis results of psychological factors about fans' aggressive behaviour.

Psychological factors	Number	Proportion (%)
Modern social pressures	583	32.12
An increase in fan appreciation	220	12.12
Management factors	174	9.59
Moral attribution and player quality	152	8.37
Corruption in the racecourse	142	7.82
The negative role of the media	126	6.94
History of the relationship between the teams in the competition	112	6.17
Nature of the competition	95	5.23
Cultural bias and inadequate punishment	71	3.91
The wrong cultural view of football	63	3.47
The sport of football itself and the characteristics of its playing process	44	2.42
Location of the competition	21	1.16
Others	12	0.66

of fans." To better analyse the causal factors of football spectator violence, a cluster analysis was conducted on the occupational distribution of this large fan community, as shown in Table 6.

In general, the fan groups in the large fan communities surveyed have the following distinctive features: (1) male-to-female ratio: men predominate, with a male-to-female ratio of 2.28 : 1; (2) age distribution: the highest proportion of people aged between 19 and 30, at 68.3%; (3) education level: mainly secondary education; (4) political outlook: mainly members of the Communist Youth League and the masses; (5) marital status: the proportion of unmarried people is higher than that of married people, with a lower divorce rate; (6) income level: the proportion of unmarried people is higher than that of married people (81.7%); (5) marital status: the proportion of unmarried people is higher than that of married people, and the divorce rate is lower; (6) income level: mainly in the range of less than 3,000 yuan (81.7%); and (7) occupational distribution: mainly in the category of "enterprise employees" (36%).

5.3. Analysis of Psychological Factors. A summary of the triggers for violence by members of the fan community, based on the first 3 main psychological factors derived from the improved K-means clustering algorithm, is as follows:

(1) Modern Social Pressure. The rapid development of all aspects of modern society, the accelerated pace of life, and increasingly fierce competition have put people under enormous psychological pressure. In the context of everyday life and work, this suppressed emotion cannot be released due to the constraints of self-imposed social identity. Depersonalization is the loss of one's identity in a crowd, and depersonalized

Table 6: Occupational distribution of the survey sample.

Occupational distribution	Proportion (%)
Business managers	9.1
Corporate staff	36
Science, education, culture, and health professionals	7.1
Professional and technical staff	5.3
General civil servants	2.4
Financial practitioners	4.3
Service industry practitioners	8.5
Individuals, private owners	4.7
Freelancers	13.7
Retired or unemployed	6.9
Others	2.01

individuals are prone to violent behaviour. Because emotions dominate thinking during the game, fans temporarily abandon the norms of everyday life and no longer care about their identity, giving rise to depersonalized and arbitrary expressions of their feelings. Fans will behave more boldly than usual and are prone to violent behaviour.

(2) An Increase in Fan Appreciation. With the rapid development of the economy, fans are also able to see higher-level matches through media such as TV or newspapers, and at the same time, the level of fans' appreciation of football is increasing. Fans are also demanding more and more from the game. If the level of excitement that can be achieved on the field does not meet the growing level of appreciation, it can lead to discontent among spectators. Fans may provoke the referee and players.

(3) Management Factors. Poor organizational management and improper maintenance of order on the field of play can also be a trigger for violence. Inappropriate enforcement tactics by match enforcers are directly related to fan violence. The failure of managers to disperse crowds assembled at the end of matches and the inability of live enforcers to stop some excesses in time can easily lead to spectator violence.

6. Conclusions

This study proposes a method for mining the psychological factors of fan community members based on improved K-means clustering. The proposed method can mine the data from a large number of psychological questionnaires to find out which factors induce fan aggression, so that effective prevention and control can be carried out. A PCA-based K-means initial clustering centre optimization algorithm is proposed for the data characteristics of the interaction of psychological factors, so that multiple causes of fan aggressive behaviour can be mined more efficiently, and the proportion of each cause can be obtained. Finally, a summary analysis of the psychological factors that lead to the generation of fan violence is made. Although the PCA-based K-means initial cluster centre optimization algorithm can reduce the number of iterations, the time complexity in the initial cluster centre selection stage is high, and therefore, the overall time is still relatively high. Subsequent research will attempt linear discriminant analysis and kernel principal component analysis as an alternative to PCA.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding this study.

References

[1] D. C. Viano, C. Withnall, and D. Halstead, "Correction to: impact performance of modern football helmets," *Annals of Biomedical Engineering*, vol. 46, no. 1, pp. 208–210, 2018.

[2] H. Sarmento, M. T. Anguera, A. Pereira, and D. Araújo, "Talent identification and development in male football: a systematic review," *Sports Medicine*, vol. 48, no. 4, pp. 907–931, 2018.

[3] Constantinou and C. Anthony, "Dolores: a model that predicts football match outcomes from all over the world," *Machine Learning*, vol. 108, no. 1, pp. 49–75, 2018.

[4] S. N. Jensen, A. Ivarsson, J. Fallby, S. Dankers, and A. M. Elbe, "Depression in Danish and Swedish elite football players and its relation to perfectionism and anxiety," *Psychology of Sport and Exercise*, vol. 36, pp. 147–155, 2018.

[5] C. Eirale, "Hamstring injuries are increasing in men's professional football: every cloud has a silver lining?" *British Journal of Sports Medicine*, vol. 52, no. 23, p. 1489, 2018.

[6] J. M. Buldú, J. Busquets, J. H. Martínez et al., "Using network science to analyse football passing networks: dynamics, space, time, and the multilayer nature of the game," *Frontiers in Psychology*, vol. 9, pp. 115–124, 2018.

[7] S. Muniroglu and E. Subak, "A modified t-test for football referees to test agility, quickness and sprint performances," *Journal of Education and Training Studies*, vol. 6, no. 5, pp. 10–17, 2018.

[8] J. P. Berkowitz and C. A. Depken, "A rational asymmetric reaction to news: evidence from English football clubs," *Review of Quantitative Finance and Accounting*, vol. 51, no. 2, pp. 347–374, 2018.

[9] H. Folgado, B. Gonçalves, and J. Sampaio, "Positional synchronization affects physical and physiological responses to preseason in professional football (soccer)," *Research in Sports Medicine*, vol. 26, no. 1, pp. 51–63, 2018.

[10] M. C. Leitner and F. Richlan, "Analysis system for emotional behavior in football (aseb-f): matches of fc red bull salzburg without supporters during the covid-19 pandemic," *Humanities and Social Sciences Communications*, vol. 8, no. 1, pp. 14–22, 2021.

[11] Y. Nagashima and A. Horikawa, "Association between 24 h urinary sodium and potassium excretion and dietary intake in Japanese male adolescent football players," *Adolescents*, vol. 1, no. 4, pp. 461–472, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9869350, 1 page
https://doi.org/10.1155/2023/9869350

Retraction

Retracted: Characteristics Analysis of Applied Mathematics in Colleges and Universities Based on Big Data Mining Algorithm Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wang, F. Tian, and Y. Bai, "Characteristics Analysis of Applied Mathematics in Colleges and Universities Based on Big Data Mining Algorithm Model," *Security and Communication Networks*, vol. 2022, Article ID 7978031, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7978031, 13 pages
https://doi.org/10.1155/2022/7978031

Research Article

Characteristics Analysis of Applied Mathematics in Colleges and Universities Based on Big Data Mining Algorithm Model

Yuqin Wang [1], Feng Tian [2] and Yong Bai [3]

[1] *College of General Education, Chongqing Vocational and Technical University of Mechatronics, Bishan, Chongqing 402760, China*
[2] *College of Mechanical Engineering, Chongqing Vocational and Technical University of Mechatronics, Bishan, Chongqing 402760, China*
[3] *Information Engineering Institute, Chongqing Vocational and Technical University of Mechatronics, Bishan, Chongqing 402760, China*

Correspondence should be addressed to Yuqin Wang; 202001277@yangtzeu.edu.cn

Received 15 February 2022; Revised 9 March 2022; Accepted 15 March 2022; Published 13 May 2022

Academic Editor: Chin-Ling Chen

To analyze the mathematical nature of applied mathematics in colleges and universities, a method based on a model of big data mining algorithms is proposed. Firstly, the modeling is carried out through the deployment of nodes, which can accurately collect the characteristics of data information in the case of massive big data; secondly, the acquisition algorithm of multi-feature fusion is systematically optimized, which can avoid data interference and collect features quickly and accurately; thirdly, by transforming the multidimensional application-oriented university applied mathematics discipline model into an unlimited experience loss minimization problem with penalty factors, the improved support vector machine algorithm is used to construct and solve the objective kernel function. It is proved that the intelligent collection method based on big data mining algorithm model for the characteristics of applied mathematics in colleges and universities is effective. The sympathetic set corresponding to the flowing data is [115~135]; the data similarity in big data environment is 1; in order to ensure that the intelligent collection method of applied mathematics discipline characteristics in colleges and universities based on big data analysis can collect data characteristics more accurately, which are 110/76.65/78/110, respectively.

1. Introduction

With the revolution of science and technology, artificial intelligence, Internet plus, and other emerging technology industries have occupied all levels of social production and life, and digital economy has changed the way of production and life of mankind. Facing the torrent of big data and the emergence of many diversified data, big data has brought opportunities and influence to personal life, enterprise management, education, and teaching, and even the development of the country and society. China's big data era has come. In the next few years, talents who can meet the development needs of big data industry will get more and better development space. However, a grim reality is before us. How to develop big data talent is an urgent issue.

Therefore, the undergraduate talent training of mathematics and applied mathematics should seize the opportunity of the times, comply with the trend of the times, clarify the guidelines and specific objectives of the party and the state, and accelerate the teaching, experiment, and talent training, and big data technology will contribute to the development of scientific research, social services, and big data industry talent, and will be the driving force behind the development of the big data industry.

Big data mining algorithm is the frontier field of current informatization. Its great power of technological change and the magic charm in business practice have attracted the favor of many experts and scholars in the field [1]. The research shows that the collection, analysis, management, and application of big data combined with high and new technologies such as data mining are not only the key to improve the development and utilization of data but also a feasible way to improve the intelligence and dynamics of human data processing, as shown in Figure 1.

2. Literature Review

With the advent of the digital age and the rising frequency of Internet use, data feature collection and extraction in massive data can effectively promote the rapid identification and efficient utilization of data. Data in the Internet are effectively transmitted in the form of data compression, and small data are also effectively transmitted in the form of multiple compressed into large data packets. Therefore, effective analysis and use of big data are the keys to use data in the future [2]. In a large data environment, the traditional method is a two-line polynomial method that breaks down large data into efficient data properties. This method can accurately extract the data features, but due to the increase of the amount of data, the traditional methods cannot be applied to the modern data environment. In the age of big data, applied mathematics has become a basic skill and a necessary tool for survival in the digital economy era and has created tremendous productivity in managing the real world through the virtual world. High technology is the application of mathematics. Exploring the accurate teaching and training of applied mathematics under the background of big data has become an important research topic.

Gao et al. believed that the construction project of applied mathematics discipline is discussed from the aspects of organizational structure, existing problems, and project management mechanism, and some suggestions are given [3]. Lu and Zhou started with the fragmentation characteristics of applied mathematics discipline construction projects in colleges and universities, and this paper analyzes all aspects of the fragmentation of applied mathematics discipline construction projects and gives three crack paths from the reform of management system, innovation of construction ideas, and specific paths [4]. Yang et al. studied the management mode of applied mathematics discipline construction project in colleges and universities with system engineering methodology, and put forward project management suggestions from four dimensions [5]. Zhang and Wang studied the objective integrated management elements and modes of mathematics discipline construction projects [6]. Ying and Zhao comprehensively discussed the project management of discipline construction, performance evaluation, target setting, and evaluation system [7]. Dbicka-Kumela et al. improved the construction project management of key disciplines by introducing the third-party supervision system and put forward corresponding suggestions [8]. Chen et al. studied the postevaluation method of applied mathematics discipline construction project by using fuzzy analytic hierarchy process, and gave the corresponding three-level evaluation index system [9]. Hui Ping and Song believed that in the data age, the most precious information in mathematics is massive educational data, which is the core cornerstone of the development of smart education. Data mining technology and learning analysis technology are the links between big data and smart education [10]. Cao et al. believed that technology's participation in the field of intelligent learning is mainly reflected in its boosting effect on learning activities, learning experience, and personalized learning [11]. Zuo and Chen believed that the intelligent learning environment should be able to detect the learning situation and distinguish the characteristics of mathematics [12]. Zhang et al. believed that the core of intelligent learning is information analysis and learning support, and information analysis plays a supporting role in determining the goal of intelligent learning [13]. Dongpo and Hongxia said in the age of the information economy, and the ability to use data plays an important role in the scientific decision-making process of enterprise development [14].

Based on the current research, this paper puts forward an analysis and processing method with the help of big data mining algorithm model, establishes the effect prediction model of applied mathematics discipline construction project management, discusses the discipline construction method based on big data mining analysis model, establishes the model, and makes prediction and analysis through the collection, sorting, and analysis of data related to mathematics discipline characteristics.

3. Characteristics of Applied Mathematics in Colleges and Universities Based on Big Data Mining Algorithm Model

3.1. Big Data Mining Algorithm

3.1.1. Concept and Characteristics of Data Mining. (1). Big Data Concept. Big data refers to a diversified, time-effective, collected, and huge data collection, which cannot be analyzed and sorted by conventional tools. Since the 21st century, the rapid development of information resources has supported the advancement of information technology and marked the arrival of the big information age [15]. The investigation and research show that the fields involved in big data include astronomy, biology, computer, electronic technology, automation, information resource management, and other aspects. It can sort out and classify according to the users' usual browsing content and concerned information, and accurately provide customers with satisfactory services. Through the analysis and sorting of big data, traditional enterprises facing Internet pressure can ensure that their products keep pace with the times.

(2). Data Mining Concepts. Data mining is the core of big data. It is not only the inevitable product of the development of the times, but also an independent emerging subject. Through data research, it is found that data mining is closely related to business planning. For potentially important information, data mining collects and arranges complex, cumbersome, and large amounts of data, which can promote the development and innovation of business [16]. At present, data mining has been applied in many fields such as education, scientific research, electronic and mechanical automation, marketing, and Internet, and has created huge

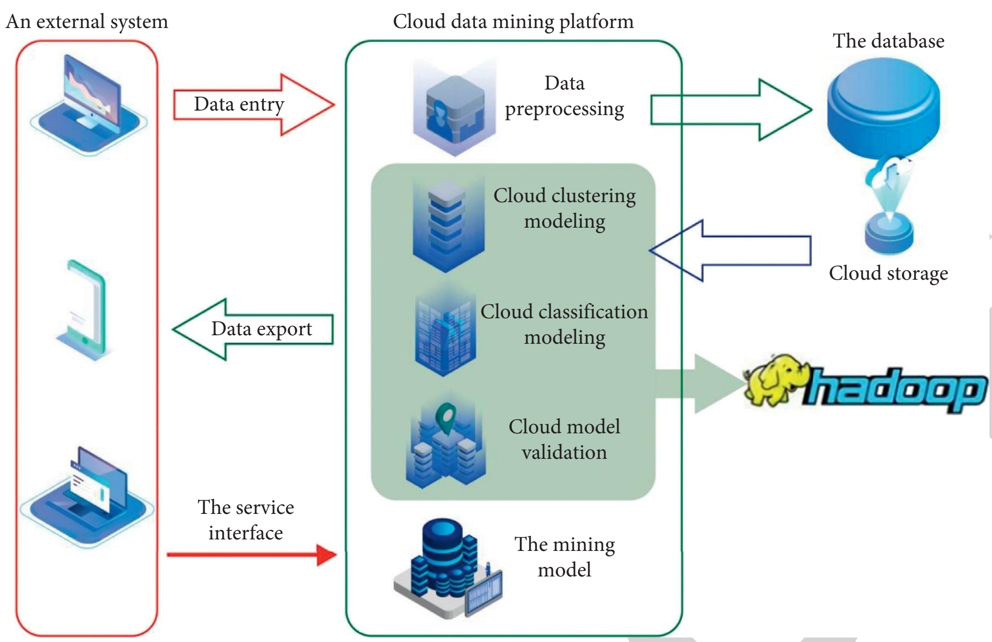


Figure 1: Cloud data mining platform.

economic benefits in many fields and promoted the rapid development of the industry.

(3). Characteristics and Methods of Data Mining. Data mining is a tool for studying valuable information from a vast body of information, and it is called information retrieval. Data mining is the process of automatically searching for hidden information of a specific relationship within a large amount of data. It collects and organizes valuable data mainly through statistics, online analysis, data retrieval, machine training, and expert systems. Data mining is considered an important step in the process of acquiring knowledge in the field of artificial intelligence. [17]. Computer technology analyzes data, finds laws from large amounts of data, and integrates new sources of information from relevant data. Data extraction includes association analysis, cluster analysis, specific group analysis, and evolution analysis. However, not all information retrieval and sorting processes are data mining. For example, database management system or Internet search engines are tasks in the field of information retrieval. This data processing includes cumbersome algorithms and precise logic. Data mining seems to be involved in a wide range of fields, but its practical application has not been fully popularized. The main development trends of data mining are further study the methods of information collection, standardize the commercial application of data mining, and establish a new system to adapt to social development.

3.1.2. Big Data Analysis Framework. The data mining algorithm designed in this paper includes clustering analysis algorithm and BP neural network model algorithm, which are organically combined to realize the analysis of various big data [18]. The data extraction algorithms developed in this article can analyze the information and knowledge processes that are valuable, used, and analyzed that are hidden in many incomplete, impure, and randomly aggregated data. How to effectively use this information and extract useful information resources from these data has become an urgent key step in the process of knowledge discovery. The architecture diagram of data processing is shown in Figure 2.

In the database, big data information not only includes the data transmitted by various sensors such as vibration sensor, temperature sensor, humidity sensor, and magnetic field sensor but also includes text, audio, video, graphics, images, and other data used in various occasions. When analyzing these data, users can select data samples for these data and extract sample features from these data samples [19]. Then, the extracted sample features are stored in the database. In this design, the data storage layer can store the result data obtained by the data acquisition layer in the form of structured database. The data mining layer analyzes different structures of data. The data mining layer provides various data mining algorithms, effectively analyzes different data types, outputs users' expected values, mines the required information for users, and effectively completes various data mining tasks.

3.1.3. Types of Data Mining Algorithms. Data mining includes many kinds, such as association algorithm, decision tree, neural network model, regression analysis, genetic algorithm, and other algorithms. Different data analysis types can realize different data output and meet the different needs of users [20]. This paper combines K-means clustering algorithm and BP neural network algorithm model to realize the analysis of big data. As shown in Figure 3, firstly, the classification of different data is realized through the cluster analysis algorithm model. Cluster analysis is a statistical analysis technology that divides a group of research objects into relatively qualitative clusters. Through cluster analysis,

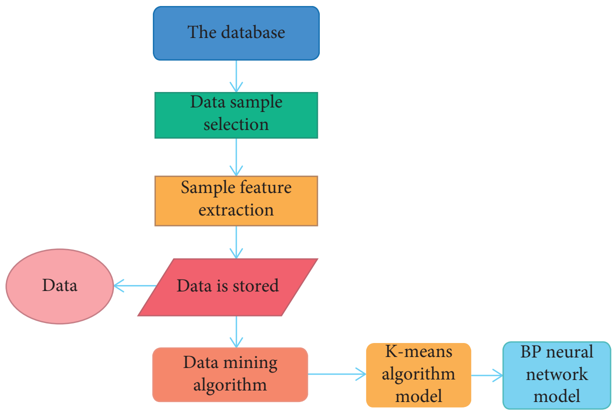


Figure 2: Big data analysis architecture design.

the data with the same attributes are divided together, which is helpful for users to realize data analysis. Then, the data after cluster analysis are input into the BP network algorithm model for further training and learning [21].

(1). Cluster Analysis Algorithm. In the design of this paper, firstly, K-means algorithm is used to preliminarily classify and calculate the data. K-means algorithm is also known as k-means clustering algorithm and fast clustering method. In this method, K-means algorithm can classify various data into a predetermined number of classes K based on the minimization error function; that is, the selected sample data are divided into k parts. The algorithm can process batch data [22]. The main steps of big data analysis are as follows:

(1) K center points are randomly selected from different sample data to select the center point of the initial cluster; before extracting data, it is preferred to preprocess the data to obtain relatively pure data, and then proceed to the next step based on the processed data. Then, among the data mentioned above, such as sensor data, power grid data, text, audio, video, graphics, and images, K sample data to be analyzed are randomly extracted. These extracted sample data can be used as the center of the initial value big data set, also known as the center of the sample cluster [23].

(2) Each different data point is configured to the center point closest to the data point, the sample cluster points are divided, and the points of multiple different sample data clusters are divided into the sample cluster represented by the center closer to it; that is, the center points closer to the center point of the initial cluster are divided into one class. In this step, the distance formula (1) is introduced to obtain

$$\mathrm{d}(x, y)^2 = \sum_{i=1}^{n} (x_i - y_i)^2 = \|x - y\|_2^2. \tag{1}$$

In formula (1), XY represents heterogeneous samples, respectively; N represents the dimension of the data sample; and DXY is the Euclidean distance. Based on the center point of the cluster sample of

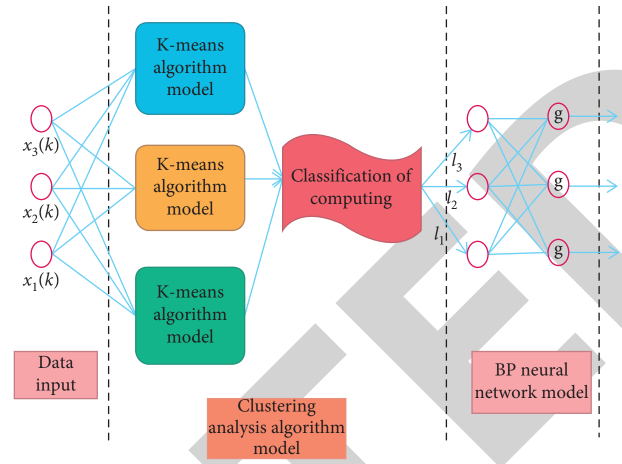


Figure 3: Data mining algorithm flow.

each data selected, the distance between each database sample data and these center sample parameters is calculated by formula (1), and the corresponding big data are redivided according to the minimum distance.

(3) Calculate the distance between each feature and these central features by the mean of the sample data object for each cluster, divide the corresponding objects by the minimum distance, and recalculate the mean of the interval distance from each point to the center. Take the point of this class and assign each data to the nearest center point. The center point of each sample data point in the sample cluster is used to represent the center point of the sample cluster. According to different parameter data, the distance between each sample data center point and these cluster information data centers can be calculated again according to the center points of different cluster information sample data, and the corresponding fault information sample data can be divided again according to the above minimum distance. Form the minimum data calculated each time into matrix D, as shown in the following formula:

$$D = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & & \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{pmatrix}. \tag{2}$$

where x is the set of minimum values.

(5) Judge whether to perform iterative calculation until all big data values are no longer allocated or the maximum number of iterations has been reached. If the number of iterations is equal to the set threshold, iterative calculation is not required. If the number of iterations is different from the set threshold, adjust the parameters and return to step 2). Repeat steps 2) and 3) to continue the calculation. In big data processing, the cluster with the smallest big data

error criterion function can be obtained by using k-means algorithm. K-means clustering algorithm is widely used and can deal with larger data sets than hierarchical clustering. Cluster K sample data points in the space as the center, and finally classify the information big data closest to different samples [24]. Through continuous iterative calculation, the values of each cluster center are gradually updated until the result of the best cluster is output, as shown in Figure 4.

(2). BP Neural Network Algorithm Model. After k-means algorithm is adopted to output the results, BP network algorithm is adopted to continue the mechanical learning and training of model algorithm, which can map and process the complex nonlinear relationship in data samples in time [25]. Because the BP network algorithm model has high learning efficiency, fast diagnosis speed, and high accuracy, it can quickly diagnose the fault type of communication data in the clustered data, making the information of processing communication fault more accurate, as shown in Figure 5.

BP neural network model includes input layer, implication layer, and output layer. In the input layer, it usually includes data such as sensor data, power grid data, text, audio, video, graphics, and images by constantly adjusting the weights and thresholds in the neural network to gradually approach the desired results and finally minimize the output error. Adjust the BP neural network model according to the following formula: the formula of the adjustment of the output layer weight system is as follows:

$$\Delta\omega_{KI} = \eta O_k^p\left(1 - O_k^p\right)\left(t_k^p - O_k^p\right)O_i^p 9. \tag{3}$$

The formula of the adjustment of weight coefficient of hidden layer is as follows:

$$\Delta\omega_{ij} = \eta O_i^p\left(1 - O_i^p\right)\sum_{i=1}^{L}\Delta\omega_{ki}O_j^p. \tag{4}$$

The formula of the quadratic accurate function model for input mode pairs in different big data samples is as follows:

$$J_p = \frac{1}{2}\sum_{k=1}^{L}\left(t_k^p - O_k^p\right)^2, \tag{5}$$

where y_i is the output database sample data information; y_i' is big data of standardized database samples; $y_{\max}, y_{\min}$ are the maximum and minimum values of the sample big data of the output database; where $0 < Q < 3$; $0 < B < 2$, then determine the number of hidden layer nodes between 6 and 10, the value from input layer to hidden layer is between 0.2 and 0.5, and the value from hidden layer to output layer can be between 0.2 and 0.3. According to the above formula, the BP neural network model can be established.

3.2. Characteristics of Applied Mathematics in Big Data Mining. Mathematics is the basis and tool of all scientific and technological development. It serves to guide and improve data mining. This can increase the efficiency of extracting valuable information from large amounts of data.

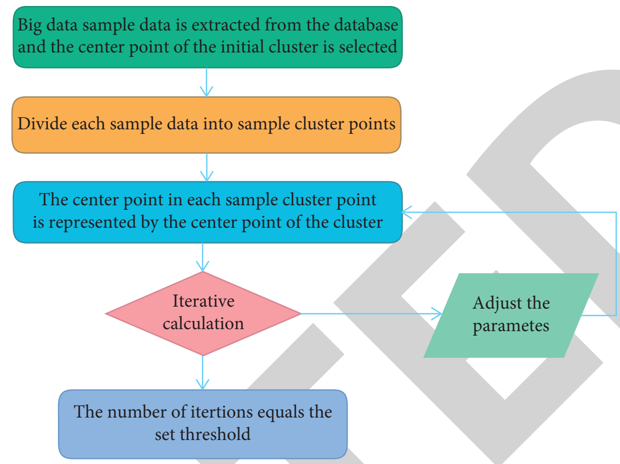


Figure 4: Flow chart of clustering analysis algorithm.

The use of mathematical knowledge in the analysis and classification of data mining will enhance the level of data analysis and stimulate the development of data mining. In the process of data mining, it is inextricably linked to the support of basic mathematical knowledge. Mathematics is closely related to big data mining. Through the application of objective function fuzzy clustering method, interval algorithm, and gray correlation analysis in data mining, this paper deeply explores the relationship between mathematics and data mining.

3.2.1. Objective Function Fuzzy Clustering Method. In big data mining, objective function fuzzy clustering method is widely used in data analysis and image processing. At present, objective function fuzzy clustering method has become the mainstream method of big data mining. This method clusters and integrates all elements through fuzzy relations through the relationship and similarity between objective things, and re-establishes the database for analysis and research. In the application of objective function fuzzy clustering method, data mining technicians use fuzzy relations to formulate certain standards for the required data, use scientific calculation methods to integrate, enrich and improve the matrix structure of data, finally collect the required clusters through direct and fuzzy clustering methods, and sort out these clusters in combination with networking method and maximum tree clustering method.

3.2.2. Interval Algorithm. Interval algorithm is a clustering method that uses mathematical means to analyze and sort out the relationship between data, and obtains important information by locking the interval value of data. Interval algorithm can integrate, mine, and process incomplete system information in the process of big data mining. Big data mining technicians use interval algorithm to convert the data encountered in the process of data mining into comparable data, and use scientific methods to sort out and analyze the data in a fixed range. Through practical research, researchers found that interval algorithm mainly includes matrix and interval clustering method, interval and interval

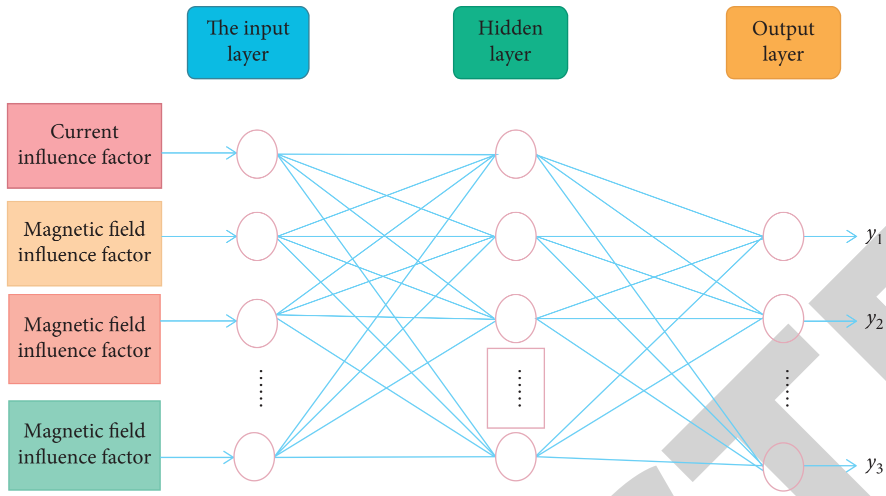


FIGURE 5: BP neural network model algorithm model.

clustering method, number, and interval clustering method. Among them, the number and interval clustering method is used most frequently. It can reasonably assist the staff to extract the incomplete system information quickly, efficiently, and accurately according to the scientific algorithm. In a clear range of values, use the most advanced statistical means and methods for scientific proof. A series of analysis and integration can also be carried out in each interval to analyze and judge the interval range of important information through practical accumulation.

3.2.3. Gray Correlation Analysis. Gray correlation analysis is a data processing method that uses the basic theoretical knowledge of gray system and measures the relationship of valuable data between large databases according to the correlation between system factors, that is, "gray correlation." This method is suitable for dynamically developing data information. In the gray correlation system, its expression is $s = (x, R)$, X is the set of influencing factors between data, and R is the set involving the mapping between factors. Big data mining technicians most often use the gray correlation analysis method to analyze the geometry between a series of messy geometric curves by scientific means and analyze the data. The more similar the shapes between geometric figures, the higher the degree of correlation between them. In data mining, the gray correlation analysis method can analyze and sort out the incomplete data and samples with less data, and get valuable information.

4. Design Scheme and Experimental Analysis of Discipline Characteristics of Applied Mathematics in Colleges and Universities Based on Big Data Mining Algorithm Model

4.1. Design Scheme of Intelligent Acquisition Method for Discipline Characteristics of Applied Mathematics

4.1.1. Establish Node Deployment Model. The intelligent acquisition method of applied mathematics subject characteristics analysis proposed uses the node deployment model to collect the system data characteristics according to the distribution of data nodes and attribute relationship. The data acquisition flow chart is shown in Figure 6.

The established node deployment model first needs to classify and reselect data, to ensure that the established node deployment model can collect data more accurately. The data classification process is as shown in the following formula:

$$\mathrm{Sim}(I, I_1) = \frac{\sum_{K\epsilon U}(P_{ik} - P)(P - P_J)}{\sqrt{\sum_{K\epsilon U}(P_{ik} - P)^2}\bullet\sqrt{\sum_{K\epsilon U}(P_{ik} - P_J)^2}}, \qquad (6)$$

where $\mathrm{Sim}(I, I_1)$ represents the data similarity in the big data environment, and the data classification is completed according to the similarity; P_{ik} represents the sympathetic set corresponding to the flow data; P and P_J, respectively, represent the preference of massive raw big data.

After the system is classified according to the similarity of the original big data calculated above, the orderly modeling can be carried out. The modeling process is completed according to the deployment of data nodes. The process is shown in the following formula:

$$p = p_1 + \frac{\sum_{k\epsilon u}\mathrm{Sim}(i,k)(p_{ku} - p)}{\sum_{k\epsilon u}\mathrm{Sim}(i,k)}, \qquad (7)$$

where p represents the established node deployment model and p_1 represents the node deployment of a kind of data. Coefficient matching is carried out according to different data. Generally, it is between [1.75–5.50] data sets; represents the characteristic representation of the original big data; and represents the attribute feature representation of big data. The features extracted from different data are stored in the set matrix column in the form of matrix. The generated matrix column is shown in Table 1.

Modeling through node deployment can accurately collect the characteristics of data information in the case of massive big data.

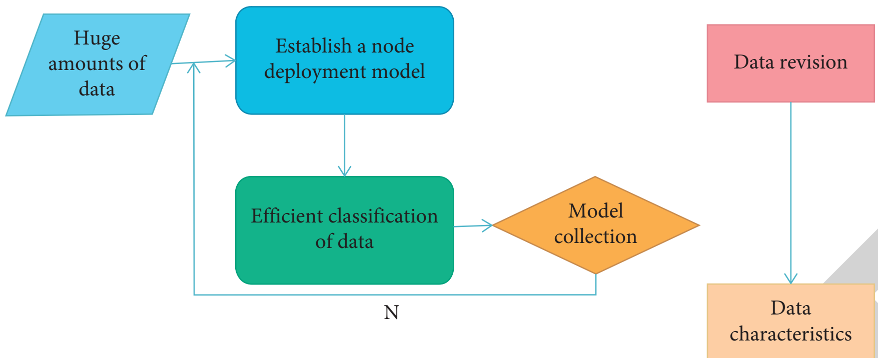


FIGURE 6: Data acquisition process.

4.1.2. Optimize the Acquisition Algorithm of Multi-Feature Fusion. Applied mathematics uses the intelligent acquisition method based on the big data analysis proposed in this paper, but uses a node placement model, but is not sufficient to collect effective data properties in the case of large data analysis. It also needs to systematically optimize the acquisition algorithm of multi-feature fusion, so as to avoid data interference and collect features quickly and accurately. First of all, the identification process of data features needs to be optimized. Formula (8) is as follows:

TABLE 1: Characteristic storage matrix column.

	I_1	I_2	...	I_N
I_1	N_{11}	N_{12}	...	N_{1M}
I_2	N_{21}	N_{22}		N_{2M}
...	...	...		...
I_N	N_{M1}	N_{M2}	...	N_{MM}

$$\operatorname{Sim}(U,U_1)=\frac{\sum_{i=1}^{n}\left(R_i-T\right)^2\times\left(R_i-T-P_{JU}\right)\times\left(R_i-T^2\times\left(R_i-T-P_{JU}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(R_i-T^2\times\left(R_i-T-P_{JU}\right)\right)_i^2+\left(R_i-T^2\times\left(R_i-T-P_{JU}\right)^2\right)^2}}, \tag{8}$$

where $\operatorname{Sim}(U,U_1)$ represents the feature attributes and feature expression set contained in big data; R_i represents the feature quantity of a certain kind of data after data classification; T represents the characteristic content; a P_{JU} represents the exclusive and unique attributes of big data. After the identification of data features, noncharacteristic attributes need to be screened out to facilitate the accuracy and speed of the acquisition process. Formula (9) is as follows:

$$P_U=\xrightarrow[R]{}+\frac{\sum_{i=1}^{n}\left(\operatorname{Sim}(U,U_1)\right)_i(R-r)_i}{\sum_{i=1}^{n}\left[\operatorname{Sim}(U,U_1)\right]_i^2}, \tag{9}$$

where P_U means to limit the screening conditions, and those meeting the conditions will be screened out; R represents the filtering conditions used in the screening process, which can prevent useful attributes from being screened out and improve the process of data feature collection; and r represents the existing screening conditions. After filtering, data features can be collected, and the process is shown in the following formula:

$$\begin{aligned}R_U&=\lambda\left(\bigcap_{i=1}^{n}r_i+\frac{\sum_{i=1}^{n}\operatorname{Sim}(U,U_1)\times(R,R_X)_i e^{i\theta}}{\sum_{i=1}^{n}\operatorname{Sim}(u,u_1)_i^2}\right),\\&+(1-\lambda)\left(R_J+\frac{\sum_{i=1}^{n}\left(\operatorname{Sim}(I,I_Y)_i-\overline{X}\right)^2\operatorname{Sim}(u,u_1)_i^2}{\sum_{i=1}^{n}\left(\operatorname{Sim}(I,I_Y)_2-\overline{X}\right)^2}\right).\end{aligned} \tag{10}$$

In the formula, $e^{i\theta}$ represents the holding weight of data features in the case of big data analysis and λ represents the balance factor correlation coefficient during acquisition. $\overline{X}$ represents the data feature acquisition factor. Through the above formula, the collection of data features can be completed, but there are certain conditions in the process of collection. The limited formula (11) is as follows:

$$\lambda=\begin{cases}\dfrac{\sum_{i=1}^{n}\phi_i^2}{\lim(M\times N)}, & M+N\geq 0,\\[2ex] \dfrac{\partial^2\Omega}{\partial u^2}\bullet M, & M+N\leq 0,\\[2ex] 0.5.\end{cases} \tag{11}$$

where $M+N$ represents the collection amount of attribute characteristics and performance characteristics in the collection process, ϕ_i^2 represents the value subattribute of big data, and $\partial^2\Omega/\partial u^2$ represents the response formula that cannot enter the acquisition process.

4.2. Implementation of University Applied Mathematics Subject Model Based on Big Data Mining. The university applied mathematics subject model based on the improved

SVM algorithm has strong adaptability. By transforming the university applied mathematics subject model for multidimensional application into an unlimited experience loss minimization problem with penalty factors, the improved support vector machine algorithm is used to construct and solve the objective kernel function. By introducing the cold and hot data separation factor and the introduction of the stochastic gradient descent factor, both space and time efficiency are achieved, and the kernel function is adaptively adjusted for different data sets. It solves the inherent disadvantages of traditional support vector machine algorithms such as poor interpretability, lack of adaptability of kernel functions, imbalance between generalization ability and learning ability, and memory consumption that changes significantly over time. Theoretically, it is applicable to the problem of high-dimensional coupled applied mathematics disciplines in colleges and universities under the constraints of any complex factors, especially for the applied mathematics disciplines in colleges and universities with single user population, no coupling of internal constraints, and large amount of collective data of interest points.

4.2.1. Improvement of SVM Algorithm. In essence, support vector machine algorithm belongs to an efficient, limited, and generalized classifier with supervised and scalable taxonomic elements, which can overcome linear and nonlinear obstacles. In order to realize the effect of nonlinear multicore data mining and clustering, improve the memory consumption ratio, balance the generalization ability and learning ability, improve the interpretability of data set, and strengthen the adaptability of kernel function. The cold and hot data separation factor and random gradient descent factor are introduced to improve the SVM algorithm. The discipline model of applied mathematics in colleges and universities for multidimensional application is transformed into an unlimited experience loss minimization problem with penalty factors. The objective function (12) of the minimization problem is defined as follows:

$$l(\omega, (x, y)) = \max\{0, 1 - y < \omega, x \gg \omega\}. \tag{12}$$

The objective function is solved by random gradient descent. In each iteration cycle, the training samples are randomly selected and the corresponding objective function gradient is reflected. The gradient step length is selected on the opposite side of the iteration direction to ensure that the running time of the algorithm meets $O(n/\lambda_c)$, where n is the sum of dimensions in the constraint space of ω hand x. In order to solve the dual problem corresponding to the nonlinear kernel, perform the following mapping transformation formula:

$$\sum \alpha_i y_i x_i \longrightarrow \omega\,(i \neq 0 \text{ and is an integer}). \tag{13}$$

After the mapping of formula (13), the unrestricted experience loss minimization problem with penalty factor under multidimensional constraints given in formula (12) is transformed into an extreme value solution problem under single constraints. Furthermore, smooth loss functions are used instead of hinge loss to further transform the problem into a smooth and unconstrained optimization problem under the hyperplane. The specific solution process is as follows:

Randomly select a training sample i_t in the hyperplane constrained space, where i represents the internal attribute of the sample and t represents the external activity of the sample (number of iterations). If it is brought into formula (12), there is formula (14)as follows:

$$f(\omega, i_t) = \frac{\lambda}{2}\|\omega\|^2 + l\left(\omega, \left(x_{i,} y_i\right)\right). \tag{14}$$

If the subgradients of formula (14) are solved, formula (15)is as follows:

$$\nabla_t = \lambda\omega_t - I\left[y_i\{\omega_t, x_i\} < 1\right] y_i x_i. \tag{15}$$

In formula (15), $I\left[y_i\{\omega_t, x_i\} < 1\right]$ is the indicator function, and the value range is two values. If it is true, it is 1, and otherwise, it is 0. Based on formula (15), we input the user interest point data set S and regularization factor λ. If the external activity of the sample (number of iterations) is T, then the iteration of one cycle can be expressed as the following formula:

$$\omega_{t+1} \leq \omega_t - \beta_t \nabla_t. \tag{16}$$

In formula (16), is the adaptive step size factor, which is negatively correlated with the number of iterations. Bring formula (15) into formula (16) to obtain the following formula:

$$\omega_{t+1} \leq \omega_t - \beta_t \lambda \omega_t - I\left[y_i\{\omega_t, x_i\} < 1\right] y_i x_i. \tag{17}$$

Further simplify formula (17) and deduce backward to obtain the following formula:

$$\omega_{t+1} \leq \left(1 - \frac{1}{t}\right)\omega_t + \beta_t I\left[y_i\{\omega_t, x_i\} < 1\right] y_i x_i. \tag{18}$$

Based on formula (18), if the indication function is true, there is formula (19)as follows:

$$\omega_{t+1} \leq \left(1 - \frac{1}{t}\right)\omega_t + \beta_t y_i x_i. \tag{19}$$

Based on formula (18), if the indicated function is not true, there is formula (20)as follows:

$$\omega_{t+1} \leq \left(1 - \frac{1}{t}\right)\omega_t. \tag{20}$$

Equation (18) shows that the cold and hot data can be separated by setting the comprehensive bias term. Furthermore, by introducing the online learning mechanism, the predictor with low generalization error can be obtained, which balances the generalization ability and learning ability.

4.2.2. Simulation Verification of Optimization Model. In order to verify the actual working effect of the applied mathematics subject model in colleges and universities after introducing the separation factor of cold and hot data and

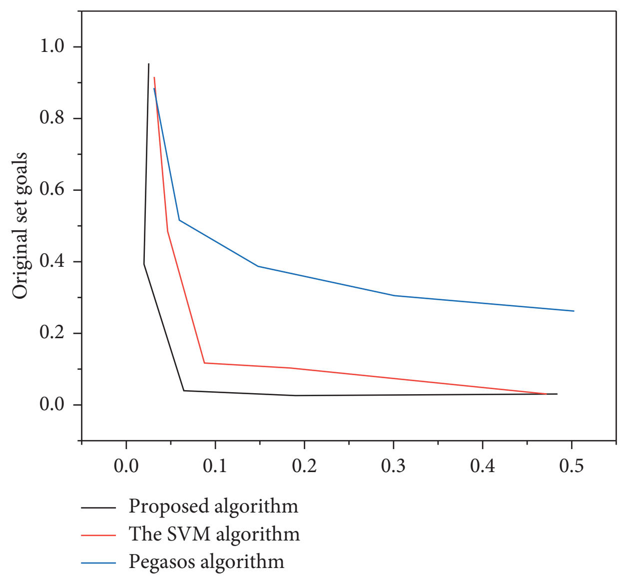


FIGURE 7: Model performance based on linear kernel data set.

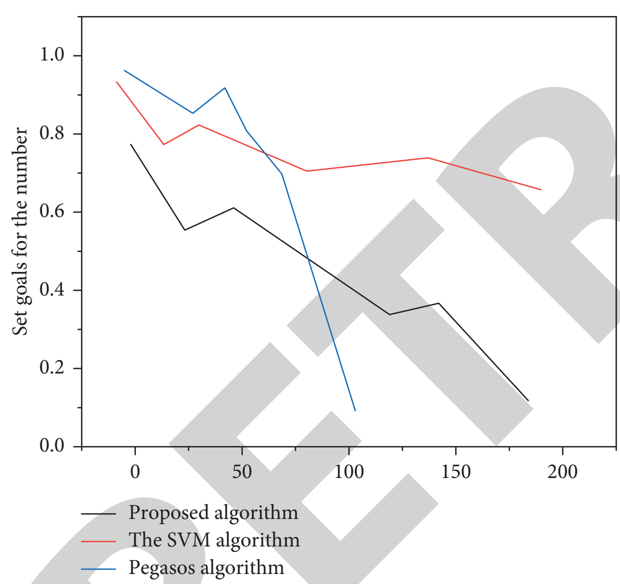


FIGURE 8: Model performance based on nonlinear kernel data set.

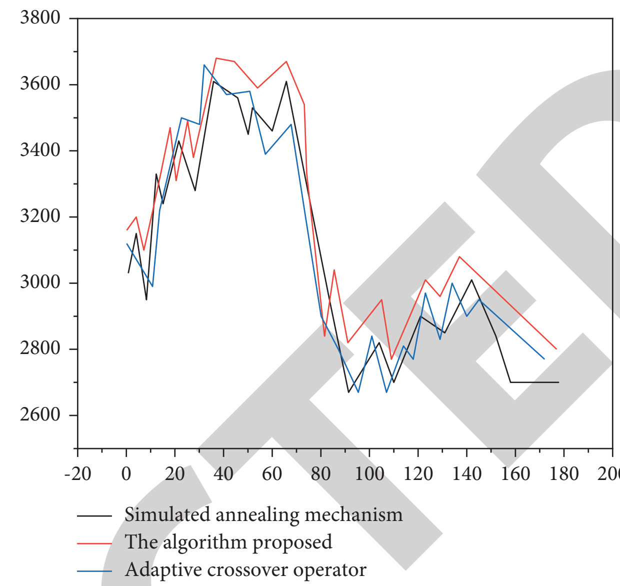


FIGURE 9: Decision analysis effect of standard fuzzy neural network algorithm.

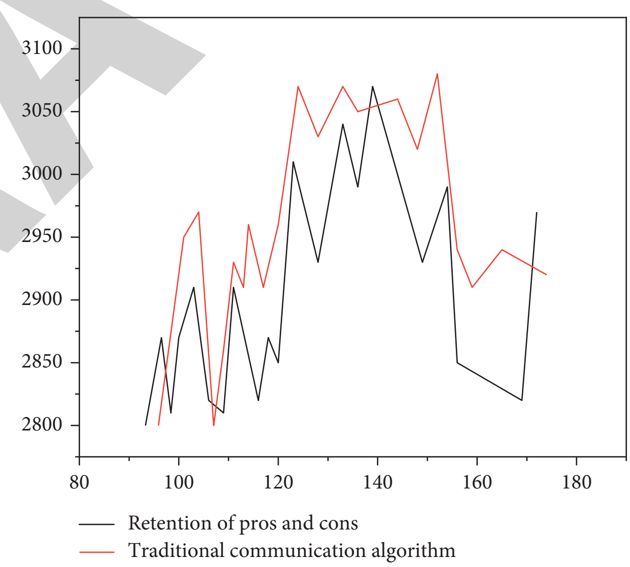


FIGURE 10: Decision analysis effect of improved fuzzy neural network algorithm.

the random gradient descent factor and integrating into the mechanism of decision analysis system, the actual efficiency of using fuzzy neural network for decision evaluation and adaptive training of fuzzy rules is analyzed. In order to be general and objective, the linear kernel data sets (ASTRO pH, CCAT, and cov1) and nonlinear kernel data sets (Reuters, adult, and USPS) are selected, set the initial objective function, and set the training times of fuzzy neural network as 1000, the training goal as 0.01, and the learning rate LR as 0.1. The algorithm is simulated and verified from multiple dimensions, such as the comparative analysis of the convergence curve of the original objective function, the comparative analysis of the convergence curve of the objective function after log regression constraint, and the comparative analysis of the classification error rate of the objective function. Based on the python 3.5.2 kernel, the graphical simulation is carried out in the PyCharm 3.5 environment. The significant difference mark is used to give the comparison curve in the simulation diagram. The final simulation results are shown in Figures 7–10, and the simulation experimental results are shown in Table 2. In order to make the test data set provided by sklearn library more appropriate to the university applied mathematics subject model, regression mapping is performed on the behavioral data set and the invisible point of interest data set, which improves the purity of the data set, reduces the data set redundancy, and improves the simulation efficiency.

Table 2: Summary of simulation experiment results.

Comparison algorithm	Linear kernel data set (Astro-ph)	Nonlinear kernel data set (Reuters)	Sample prediction accuracy (%)	Classification error probability (%)
Pegasos algorithm	Astro-ph	Reuters	66.32	50.65
This paper presents an algorithm	Dimension: 98445	Dimension: 98445	80.21	87.32
SVM-perf algorithm	Number of positive samples: 7021	Number of positive samples: 8553	75.92	56.95
LaSVM algorithm	Number of negative samples: 25627	Number of negative samples: 23562	51.97	74.15

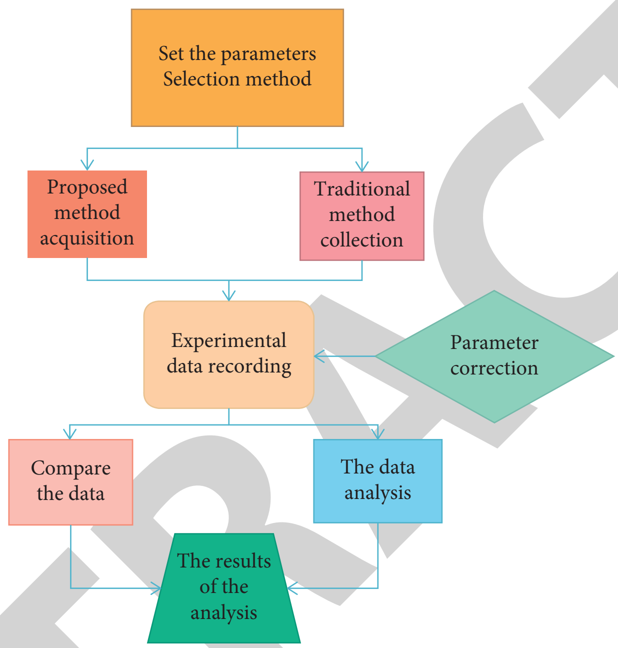


Figure 11: Experimental flow.

Table 3: Experimental parameters.

Number of experiments	Number of data features to be collected	Random feature collection	Acquisition correction parameters
1	10	5	10^{-4}
2	20	10	10^{-4}
3	30	15	10^{-4}
4	40	20	10^{-4}
5	50	25	10^{-4}

From Figures 7–10 and Table 2, it can be seen from both qualitative and quantitative aspects. Firstly, the convergence speed and fine granularity of the objective function of the improved SVM algorithm, the convergence efficiency of the objective function after log regression constraints, and the classification error rate of the objective function have been greatly improved; secondly, after introducing the separation factor of cold and hot data and the random gradient descent factor and integrating into the mechanism of decision analysis system, the actual work effect of applied mathematics discipline model in colleges and universities has been greatly improved. The actual efficiency of using fuzzy neural network for decision evaluation and self-adaptive training of fuzzy rules is good. The cooperative adaptation of linear kernel and nonlinear kernel and the consideration of space and time efficiency are preliminarily realized; finally, the kernel function is adjusted adaptively according to different data sets, which greatly reduces the memory consumption, improves the iteration efficiency of data sets, and solves the inherent disadvantages of traditional support vector machine algorithms, such as poor interpretability, lack of self-adaptability of kernel functions, imbalance between generalization ability and learning ability, and significant changes in memory consumption with time. Theoretically, it is applicable to the problem of high-dimensional coupled applied mathematics disciplines in colleges and universities under the constraints of any complex factors, especially for the applied mathematics disciplines in colleges and

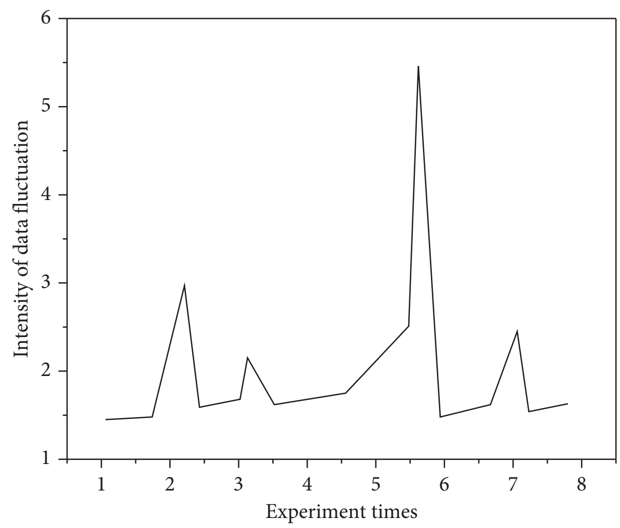


Figure 12: Data fluctuation intensity.

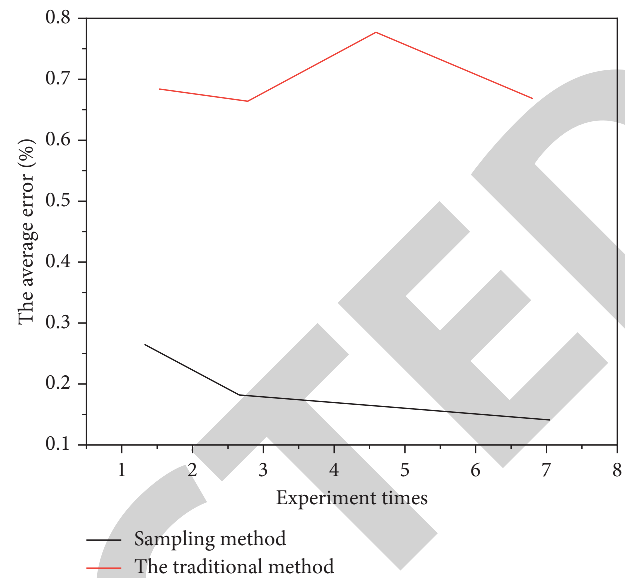


Figure 14: Comparison of experimental results.

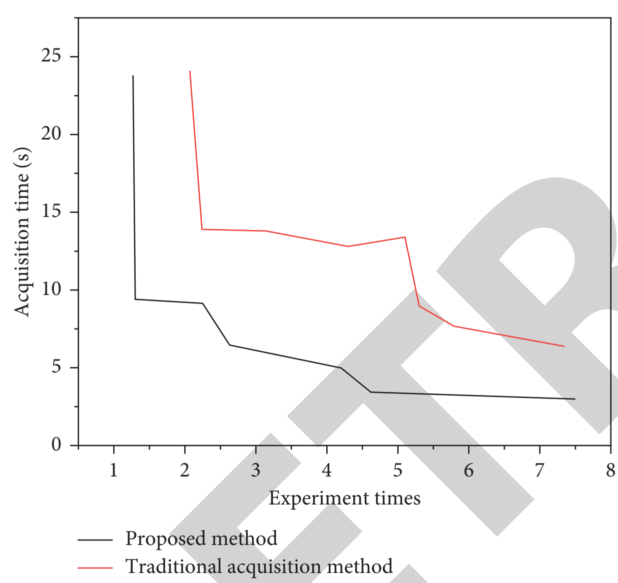


Figure 13: Comparison of experimental results.

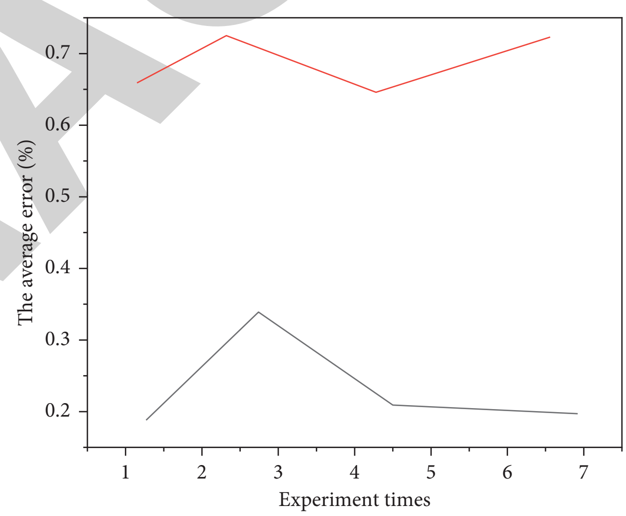


Figure 15: Comparison of experimental results.

universities with single user population, no coupling of internal constraints and large amount of collective data of interest points.

4.3. Experimental Analysis. In order to test the effectiveness of the method of intelligently mastering the characteristics of applied mathematics in colleges and universities based on big data analysis, a comparative simulation experiment is designed. Conduct in-depth feature collection of applied mathematics in colleges and universities under the environment of simulated big data. Using the intelligent acquisition method of applied mathematics discipline characteristics in colleges and universities based on big data analysis proposed in this paper. In order to ensure the effectiveness of the test, the traditional data feature acquisition method is used to carry out the test at the same time. The test process is shown in Figure 11.

In order to ensure the effectiveness of the intelligent collection method, which has developed the characteristics of applied mathematics courses in colleges and universities based on big data analysis, the parameters are set, and the sympathetic set of p_{ku} corresponding flow data is [115~135]; set the data similarity in big data environment to 1; in order to ensure that the intelligent collection method of applied mathematics discipline characteristics can collect data characteristics more accurately, $R_i, \phi_i^2, e^{i\theta}, p$ are set to 110/76.65/78/110, respectively.

The set test parameters are shown in Table 3, and the set data fluctuation is shown in Figure 12.

The experiment is carried out according to the parameters set by the above simulation, and the results are as follows.

4.4. Result Analysis. During the test, the test results of the traditional mining system and the mining system developed in this document were recorded. Analyzing the results in Figure 13, we know that the intelligent collection method based on the big data analysis proposed in this paper can collect the specific characteristics of the information in a short period of time.

By analyzing the results in Figures 14 and 15, the massive data feature intelligent acquisition method based on big data analysis proposed in this paper can maintain a low average error under the condition of obvious data fluctuation, and the trend is downward.

5. Conclusion

In conclusion, the intelligent acquisition method with mathematical characteristics of university application based on the big data analysis proposed in this paper ensures the accuracy of the data acquisition process in a large data environment by creating a node placement model, and can quickly and accurately collect the features of massive data in the big data analysis environment. The acquisition algorithm of multi-feature fusion is optimized to improve the sensitivity to data features and ensure the speed of feature acquisition. To ensure the effectiveness of the intelligent collection method of applied mathematics discipline characteristics, the sympathetic set corresponding to flowing data is [115~135]; set the data similarity in big data environment to 1, in order to ensure that the intelligent collection method of applied mathematics discipline characteristics can collect data characteristics more accurately, which are 110/76.65/78/110, respectively.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of Chongqing (No. 501100005230).

References

[1] L. Ji, D. Dai, R. Sun, W. Sun, and T. Yu, "Research on the characteristics of college students' physical fitness database in mathematical statistics," *Journal of Physics: Conference Series*, vol. 1952, no. 4, Article ID 042047, 2021.

[2] yangzhao, "Research on the application of university teaching management evaluation system based on apriori algorithm," *Journal of Physics: Conference Series*, vol. 1883, no. 1, Article ID 012033, 2021.

[3] X. Gao, S. Liu, and F. Xu, "Research on the university fall enrollment strategies based on apriori algorithm," *Journal of Physics: Conference Series*, vol. 1651, no. 1, Article ID 012093, 2020.

[4] L. Lu and J. Zhou, "Research on mining of applied mathematics educational resources based on edge computing and data stream classification," *Mobile Information Systems*, vol. 2021, no. 7, pp. 1–8, 2021.

[5] Y. Yang, M. Ma, X. Zhu, and W. Ge, "Research on spatial characteristics of metropolis development using nighttime light data: ntl based spatial characteristics of beijing," *PLoS ONE*, vol. 15, no. 11, Article ID e0242663, 2020.

[6] P. Zhang and C. Wang, "Research on the characteristics of English text based on computer and its translation," *Journal of Physics: Conference Series*, vol. 1992, no. 3, Article ID 032006, 2021.

[7] F. Ying and J. Zhao, "Research on the cultivation of engineering cost professionals based on the obe concept of big data analysis," *Journal of Physics: Conference Series*, vol. 1744, no. 4, Article ID 042080, 2021.

[8] M. Dbicka-Kumela, B. Romanowska-Dixon, I. Karska-Basta, J. Kowal, and A. Markiewicz, "The evaluation of the malignant characteristics of conjunctival lesions based on the dermatoscopic algorithm," *Anticancer Research*, vol. 41, no. 2, pp. 895–903, 2021.

[9] H. Chen, Y. Wang, and H. Zhang, "Research on the negative resistance characteristics of silicon-based trench mos barrier Schottky diodes," *Journal of Physics: Conference Series*, vol. 1670, no. 1, Article ID 012005, 2020.

[10] L. Hui-Ping, L. Song, Phase retrieval with phaselift algorithm," *Applied Mathematics:A Journal of Chinese Universities*, vol. 35, no. 4, pp. 104–127, 2020.

[11] T. H. Cao, T. Thao, N. T. Trung, T. Le, and T. Trung, "A bibliometric review of research on stem education in asean: science mapping the literature in scopus database, 2000 to 2019," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 16, no. 10, Article ID em1889, 2020.

[12] Y. Zuo and C. Chen, "Research on the characteristics of gis-based industrial agglomeration in shenyang," in *Proceedings of the E3S Web of Conferences*, vol. 253, no. 2, Article ID 02067, Hunan, China, November 2021.

[13] Z. Zhang, S. Li, E. Wang, X. Liu, and Y. Zhang, "Research on the damage evolution characteristics of coal based on cluster analysis of temporal-spatial dimension of acoustic emission events," *Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering*, vol. 39, no. S2, pp. 3338–3347, 2020.

[14] W. Dongpo and Z. Hongxia, "Research on flow characteristics of throttle valve based on fluent," *Journal of Physics: Conference Series*, vol. 1676, no. 1, Article ID 012021, 2020.

[15] Y. Liu, "Characteristics of organization and management of private colleges and universities based on the big data analysis," *Journal of Physics: Conference Series*, vol. 1744, no. 4, Article ID 042090, 2021.

[16] S. Sun, J. Cui, and T. Du, "Research on the influence of vibrations on the dynamic characteristics of ac contactors based on energy analysis," *Energies*, vol. 13, no. 3, p. 559, 2020.

[17] W. Huang, "Research on the revolution of multidimensional learning space in the big data environment," *Complexity*, vol. 2021, no. 4, Article ID 6583491, 12 pages, 2021.

[18] Q. Yuan and J. Li, "Research on the characteristics of environmental air pollution and its countermeasures in tianjin,"

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9896209, 1 page
https://doi.org/10.1155/2023/9896209

Retraction

Retracted: A Prediction and Evaluation Model Analysis of Enterprise Economic Management Mode Based on Neural Network Strategy

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Wang and X. Han, "A Prediction and Evaluation Model Analysis of Enterprise Economic Management Mode Based on Neural Network Strategy," *Security and Communication Networks*, vol. 2022, Article ID 9639025, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9639025, 10 pages
https://doi.org/10.1155/2022/9639025

Research Article

A Prediction and Evaluation Model Analysis of Enterprise Economic Management Mode Based on Neural Network Strategy

Zijin Wang [1] and Xiujuan Han [2]

[1] *School of Economics and Management, Tiangong University, Tianjin 300387, China*
[2] *College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China*

Correspondence should be addressed to Zijin Wang; 1910650323@tiangong.edu.cn

Received 24 February 2022; Revised 8 April 2022; Accepted 18 April 2022; Published 12 May 2022

Academic Editor: Chin-Ling Chen

Intelligent finance is an inevitable product for continuous development of big data, which is also a weapon to improve the work efficiency of enterprise economic management. The significance and feasibility of an artificial intelligence technology in corporate management performance have been analyzed. Two types of neural networks are used, which one is the single BP neural network without LSTM and the other is a BP neural network with LSTM layer is used to capture corporate time characteristics of performance factors between cost-savings ratio and corporate performance factors. The results have shown that the two types of BP neural networks have good accuracy in predicting corporate performance and which the prediction errors are within 5%. And both training loss and test loss have good convergence for predicting corporate performance. However, the BP neural network with LSTM layer has better accuracy than a single BP neural network. The correlation coefficient reached 0.97, which shows that the BP neural network model established in this article has good accuracy in predicting corporate performance, which is sufficient for predicting corporate performance. The application prediction errors of BP neural network in enterprise performance are all within the acceptable range, and the maximum error is only 1.23%.

1. Introduction

Enterprise economic management is a process of continuously setting goals, checking goals, and finding countermeasures. The implementation of enterprise economic management requires the support of corresponding corporate culture [1]. Only a good performance management system cannot guarantee its effectiveness. The design of the performance management system only plays a role in platform construction. As a brand new information science technology, big data technology can sort out a large amount of valuable information from massive data, which improve the accuracy of decision-making [2]. In the era of big data, with the advancement of the management accounting system, management accounting has been paid more and more attention in corporate management, especially in corporate performance management [3]. However, there are still some problems for company performance management [4]. Not enough attention is paid to the application of management accounting in performance management, and the degree of integration between the performance management system of management accounting and the actual situation of the enterprise is not enough.

Artificial intelligence has strong application value in enterprise economic management [5]. There have been many examples of AI technology application in enterprise economic management, such as early warning of potential financial crises and assessment of financial crises. Enterprise economic management is the core content of an enterprise company [6]. The level of enterprise economic management not only can affects the financial status but also can affect the development and operation of the company. The traditional corporate enterprise economic management model generally converts the data of corporate business activities into useful accounting information in accordance with prescribed procedures and processes for the use and decision-making of relevant personnel [7]. At the same time, enterprise economic management also organizes and summarizes the business that occurs during the production and operation of the company to form a financial report [8].

For the first time, the DEA method had been used to analyze the changes trend for technical efficiency and decomposition indicators of listed companies in China's big data industry [9]. The factor analysis method is used to measure the business performance of listed big data companies and their relationship with the intensity of R&D investment [10]. The research results of scholars on the evaluation index of enterprise innovation performance, combined with the observation and analysis of the current technological innovation process of large data enterprises times, have constructed a large data enterprise innovation performance multi-index evaluation system [11]. Huang et al. discussed whether big data can bring benefits to corporate performance, and the results show that the implementation of big data by companies has a positive effect on financial performance improvement [12]. Caputo et al. believed that big data analysis methods can process all kinds of data into useful information for the business. This information is the enterprise's knowledge assets. The enterprise conducts knowledge asset management to realize the value of big data and improve enterprise performance [13]. Rehman analyzed the correlation between the application of corporate human resources big data and overall performance, which pointed out that corporate human resources departments use big data technology to manage personnel, which can improve corporate efficiency and performance [14]. Based on the AHP-DEA method, Mustafa et al. selected Bolsa to evaluate the performance of real estate trust investment funds, and which put forward relevant research conclusions that expanding scale and relying on economies of scale can improve performance [15]. Mao used the Todim method to study the application of performance evaluation in strategic emerging industries [16]. Md et al. combined fuzzy comprehensive evaluation and decision-making experiment and evaluation laboratory (DEMATEL) methods to resolve the interdependence between attributes in the corporate performance structure [17]. Tai found that there is actually a positive (negative) correlation (downgrade) between upgrades and abnormal returns when the company's financial performance is better. It further clarifies the correlation between corporate governance evaluation activities, abnormal returns, and the company's financial performance [18]. Jiang and Xue studied the impact of corporate environmental responsibility (CER) and ownership structure on the corporate performance. The results verify that CER has a positive impact on corporate financial performance [19]. Many researchers have done a lot of research on Business performance management, but the application of neural network methods in Business performance management is less studied. In this paper, the main combination of BP neural network technology is to assess the performance of enterprises.

At this stage, with the continuous expansion of the scale of enterprises, the acceleration of capital flow, and the linear increase in the frequency of information, it is obvious that the traditional financial model can no longer adapt to the current rapid economic development [20, 21]. From the past to the present, one of the biggest problems existing in the enterprise economic management system in corporate companies is the disconnection between actual management and business [22]. At the same time, due to the advent of the era of big data, the traditional enterprise economic management model has been unable to efficiently and orderly complete the processing of data and filter out useful information in a timely manner. This drawback has hindered the rapid development of enterprises to a certain extent [23]. While smart finance brings development opportunities and advantages to mankind, it also faces many challenges and opportunities at present. The integration of artificial intelligence and enterprise economic management is an emerging field. Many development trends of intelligent finance have not been clarified, and many challenges will be faced. At this stage, intelligent finance combined with a rapidly development of the artificial intelligence method has obtained the best development opportunities, but at the same time it is also facing many challenges [24]. There are many data with strong correlation in the enterprise performance model, but it is very difficult to process these numbers manually and find the correlation among them. BP neural network is a model with forward propagation and back propagation. It can fit nonlinear data and find the correlation among them. This is suitable for the enterprise performance evaluation system [25]. There are a lot of data to be crunched in the business performance management, and it is a tricky task to rely on professionals to do it alone. The advantage of BP neural network is the processing of Galway nonlinear data, it can well handle the relevant complex data in the business performance management.

This article is mainly composed of five chapters. The first section is the development status of the combination of enterprise management performance and big data. The significance of corporate performance evaluation and large data sets are studied in Section 2. The third part introduces the theory and training process and testing process of two types of BP neural network. The fourth section describes the iterative process of the loss function of the BP neural network in the training phase and the test phase. Section 4 mainly introduces the feasibility and accuracy of BP neural network in business performance management by some statistical parameters. These statistical parameters mainly include error and correlation coefficient, prediction distribution curve, etc. At the same time, in order to more intuitively reflect the accuracy of the predicted value and the true value, it is reflected by the linear correlation coefficient. Finally, the applicability and accuracy of two different types of BP neural network models are verified. Finally, summary is given in Section 5.

2. The Significance of Corporate Performance Evaluation and Large Data-Sets

2.1. The Significance of Big Data for Performance Management. Effective performance evaluation is the key to employee participation, and it can also provide valuable feedback on skills and goals that are important to the success of corporate business [26]. Enterprise performance management is a kind of evaluation by supervisors of work performance of employees. In the evaluation process, the supervisor will determine the strengths and weaknesses of the employees, which will set goals and provide predictions and feedback on future performance of employees. The enterprise performance evaluation system consists of a series of evaluation systems, evaluation index systems, evaluation methods, evaluation standards, and evaluation institutions related to performance evaluation [27]. Companies should consider whether it has laid a good foundation for an establishment of a performance evaluation system in the process of their own development and reflect on whether they have the conditions to establish a value evaluation system. When the company is in a higher stage of development, the internal soft power will slowly grow and strengthen, the corporate culture will gradually become stable, and the quality of the selected employees will often be relatively high [28]. In the era of big data, the storage, collection, and processing of corporate financial information are more convenient, which provides the possibility for the full realization of the management accounting function. Management accounting adds value to the company through data analysis and has important advantages in corporate cost management, operation management, and performance management [29]. The application prospect is broader. Enterprises should pay attention to the innovative application of management accounting in the evaluation system and evaluation content in performance management, so as to make the performance evaluation at all levels of the enterprise more real and effective [30].

First of all, the large data sets technology has changed the traditional way of performance management. The purpose of performance management is not to evaluate, but to ensure that the direction of the organization's development is correct. Therefore, the target value in performance management should be dynamic and adjusted according to changes in the environment. Performance results are not used for employee appraisal, but only for goal correction. The performance goals are set according to influencing factors, and recommendations are given according to certain algorithms [31]. The performance goals are adjusted in real time, and it also can be updated according to each performance completion of staff. The large data sets technology is used to help companies comprehensively utilize massive amounts of data and which can quickly collect, manage, process, and organize data into helpful information to help companies make business decisions. The large data sets technology includes data processing tools (such as R language, etc.), data analysis theories and methods (such as regression, clustering, Bayesian, etc.), the large data set analysis tools, and data visualization [32]. The combination of corporate performance and big data is meaningful research, which will help companies better predict future development trends.

2.2. The Preparation Process of Data-Sets. There are many sources of data, and it is necessary to select stable and testable data sources according to the needs of different indicators. However, it is generally necessary to analyze statistical calibers and statistical schemes for data discrepancies. If the discrepancies are caused by statistical means, it can be ignored. If a data source has large abnormal fluctuations, it can often use another data source for comparative analysis. If the two sides have fluctuations in the same direction and the same magnitude, then it needs to be analyzed from the performance indicators. If the two sides are very different, it is likely to be the data. Data sorting and cleaning are mainly to exclude dirty data sets and abnormal data, which are used to structure the data sets. Figure 1 shows the prediction process of BP neural network. First, the training data sets (such as employee performance, etc) are normalized, the training set is processed into normally distributed data to speed up the convergence speed more quickly, then continue to optimize the weight and bias through the back-propagation method. Once the model is trained, which the weights and biases are applied to the prediction of this research problem, and the test set can be quickly predicted. For the test set, once the BP neural network training completed the training set, the enterprise managers can pass unknown performance parameters to BP neural network, and it can be an efficient output.

3. Method and Theory

3.1. BP Neural Network. BP neural network is one of the most basic neural networks compared to deep learning methods, and it is mainly composed of fully connected layers. It has been successfully used in many fields. It also has strong nonlinear ability and dimensionality reduction ability. It is mainly composed of input layers, hidden layers, output layers, and etc., which the input layer allows multiple inputs. Weights and biases are the parameters that need to be learned. The directional propagation ways and the gradient descent ways are used to continuously fit the reversed error between the true value and the predicted value until the error function converges to find the optimal weight and bias. Figure 2 shows the structure of BP neural network. In this article, corporate performance value and employee personal performance are used as output and input of BP neural network, respectively, it can map the nonlinear relationship between input and output. In this study, a BP neural network with 4 hidden layers was selected, and the learning rate was set to 0.001.

The BP neural network also can be divided into two processes. The first process is a forward operation process. It performs matrix operations on the input, weight, and offset, and which then nonlinearizes the matrix through the activation function to obtain a certain matrix at a certain iteration step. The second step is backpropagation method, which first can calculate the predicted value and the true value through a loss function, and after determining the loss value, performs a derivative operation according to the backpropagation method and automatic differentiation technology to find the region of gradient descent. BP neural network continuously iteratively

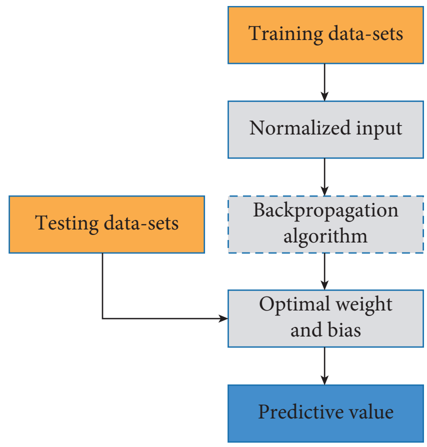


Figure 1: The predictive steps of BP neural network.

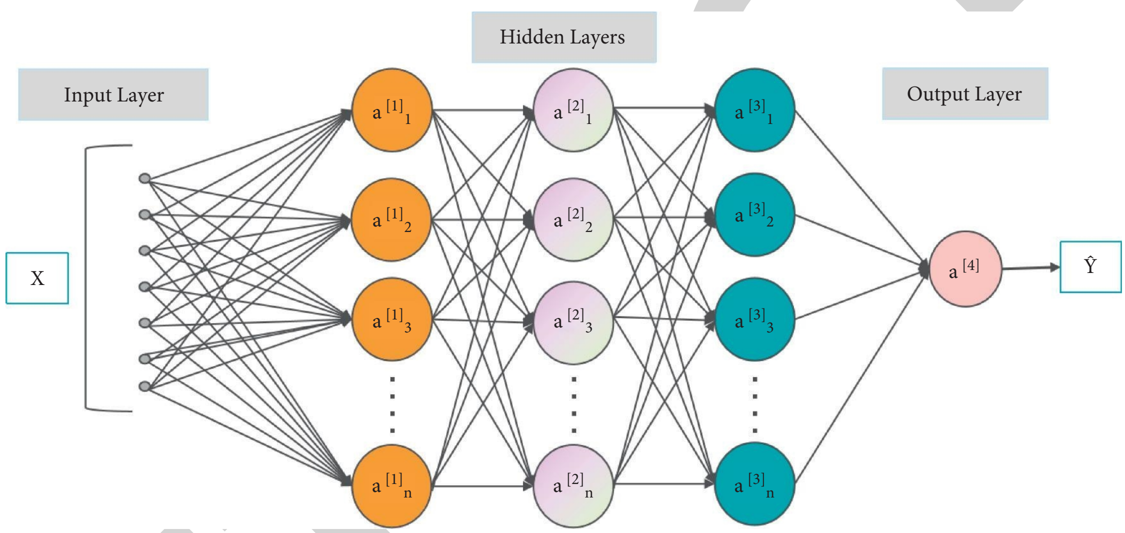


Figure 2: The structure of BP neural network.

searches for the smallest gradient according to these two steps and then finds the optimal weights and bias parameters.

The study selects a network with three hidden layers as an introduction, the difference between the predicted value and the true value is the propagation error, which is an output error E of back-propagation, E can be defined in equation (1):

$$E = \frac{1}{2}\left(d_{\text{out}} - O_{\text{real}}\right)^2 = \frac{1}{2}\sum_{\kappa=1}^{t}\left(d_\kappa - O_\kappa\right)^2, \tag{1}$$

where d is the predictive value of saving cost ratio in this study and O is a real value at every days. It can expand above error of equation (1) definition to the hidden layer and further expand to the input layer:

$$E = \frac{1}{2}\sum_{k=1}^{m}\left[d_k - f\left(\text{net}\omega_k\right)\right]^2 = \frac{1}{2}\sum_{k=1}^{m}\left[d_k - f\left(\sum_{j=0}^{n}\omega_{jk}y_j\right)\right]^2,$$

$$E = \frac{1}{2}\sum_{k=1}^{m}\left[d_k - f\left(\text{net}\omega_k\right)\right]^2 = \frac{1}{2}\sum_{k=1}^{m}\left[d_k - f\left(\sum_{j=0}^{n}\omega_{jk}y_j\right)\right]^2$$
$$= \frac{1}{2}\sum_{k=1}^{m}\left[d_k - f\left[\left(\sum_{j=0}^{n}\omega_{j\kappa}f\left(\sum_{j=0}^{q}u_{ij}\chi_i\right)\right)\right]\right]^2, \tag{2}$$

where ω_{jk} is the weight matrix. Obviously, reducing the error is actually looking for the inverse of gradient descent, and the error gradient descent method is as follows:

$$\Delta\omega_{ji} = -\eta\frac{\partial E}{\partial \omega_{ji}},$$
$$\Delta u_{ij} = -\eta\frac{\partial E\, \partial}{\partial u_{ij}}. \tag{3}$$

3.2. The LSTM Neural Network. In corporate performance evaluation, time characteristics are often also very important. Long and short memory neural networks are very suitable for extracting temporal information features. Corporate performance evaluation is not only the relationship between different influencing factors but also the relationship between the same factor at different times. Corporate economic performance management is often closely related to time. The long and short times memory neural network (LSTM) is used to extract the time characteristics of the enterprise economic data sets. Figure 3 shows the process steps of LSTM models structure. LSTM neural network has obvious advantages in dealing with temporal features, and enterprise performance management evaluation is a feature that is closely related to temporal features. LSTM is used to extract temporal features in enterprise performance management.

Compared to the single BP network without LSTM, the LSTM neural network has a structural change and has a memory function, which is mainly due to the existence of the gate structure, it is the reason why LSTM has the advantage of time memory. BP neural network is similar to convolutional neural network. It can effectively extract and map space-related features. It has strong nonlinearity, but it is difficult to learn time-related features. Due to the existence of forget gates, memory gates, and other structures, LSTM filters historical information, extracts useful historical information, and removes historical information with little relevance, thereby maintaining characteristic information with temporal characteristics. And each layer of LSTM has a strong connection, mainly to prevent useful historical information from being forgotten. The number of LSTM layers used in this study is 3, and the learning rate is 0.0001.

As shown in equation (4), the first step in LSTM is to decide what information to discard from the cell state. The forget gate acts on the LSTM state vector to control the impact of the memory of the previous time stamp on the current time stamp. Parameters such as weights and biases are solved by automatic differentiation technology. When the gate control is equal to 1, the forgetting gates are all open, and LSTM receives all the information of the previous state. When the gate control is equal to 0, the forgetting gate is closed, and LSTM directly ignores and outputs a zero vector. Where σ is the activation function, w_f is the weight matrix, and the h_{t-1} is the output value at the last moment.

$$f_t = \sigma\left(w_f \bullet [h_{t-1}, P_t] + b_f\right). \tag{4}$$

As shown in equations (5) and (6), the input gate is used to control the LSTM acceptance of historical information. First, a new input vector is obtained by nonlinear transformation of the input of the current time stamp and the output of the previous time. The input gate controls the amount of input accepted. The control variables of the input gate also come from the input and output. Tanh nonlinear function normalizes the input to between −1 and 1.

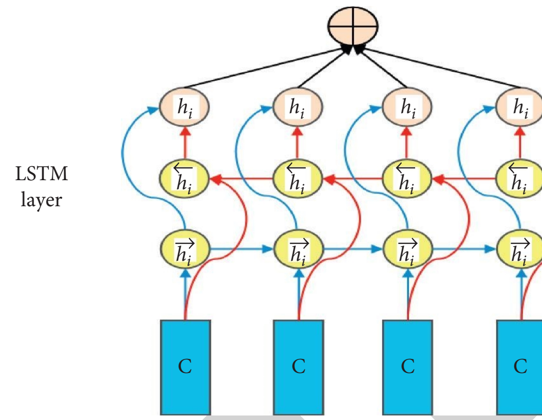


Figure 3: The application process of LSTM model in enterprise performance evaluation.

$$i_t = \sigma\left(\omega_i \bullet [h_{i-1}, P_t] + b_i\right), \tag{5}$$

$$\widetilde{C}_t = \tanh\left(w_c \bullet [h_{t-1}, P_t] + b_c\right). \tag{6}$$

After passing the forget gate and memory gate to get the current time variable, which can refresh the variable by the following formula, as shown in equation (7).

$$\overrightarrow{C}_t = f_t \times \overrightarrow{C}_{t-1} + i_t \times \overrightarrow{C}_t. \tag{7}$$

As shown in equations (8) and (9), when the output gate is equal to 0, the output is closed, and the internal memory of LSTM is completely cutoff and cannot be used as an output. When the output gate is equal to 1, the output is fully opened, and the state vector of LSTM is all output:

$$O_t = \sigma\left(w_o \bullet \left[\overrightarrow{h}_{t-1}, P_t\right] + b_o\right), \tag{8}$$

$$\overrightarrow{h}_t = O_t \times \tanh\left(\overrightarrow{C}_t\right). \tag{9}$$

3.3. Normalized Method. Because the input of enterprise performance factors is in different forms, there are certain differences in the form and magnitude of the input, which is unfavorable for the training of BP neural network, and there is a large distribution difference in the amount of input. Employee performance and labor are normalized into a data set conforming to a normal distribution, and its value remains between 0 and 1. Normalizing the input data with better distribution characteristics and correlation can speed up the convergence speed and improve the prediction accuracy. Figure 4 shows the normalized and without normalized methods. The left side of Figure 4 shows the distribution before data normalization, and the right side shows the distribution after data normalization. It can be seen that the data set has better correlation after being normalized, which is beneficial to the training process. In this study, the business performance management data were data preprocessing using a standard normalization method.

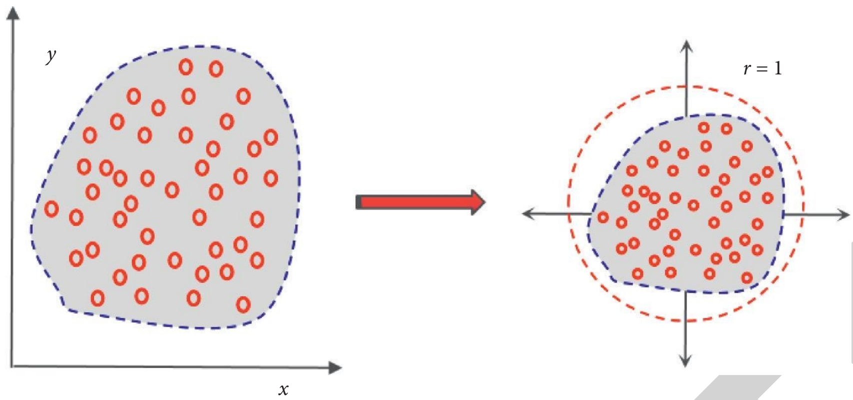


Figure 4: The normalized and without normalized methods.

3.4. Loss Function and Activation Function. In the process of forward propagation, the input, weight, and bias need to be subjected to activation function for nonlinearization after matrix operation. If the activation function is not processed, the network will lose the nonlinear ability of fitting. The activation function is the source of nonlinearity in the neural network. If the activation function is removed, then the entire network will only have linear operations. In this study, the Sigmoid function is adopt. The Sigmod function is simple to implement and the derivative is easy to obtain; its output is within the interval of [0, 1], so it can be used as the output layer to represent the probability; and it is less affected by noise data. The expression of the Sigmoid function is as shown in equation (10).

$$S(x) = \frac{1}{1 + e^{-x}}. \tag{10}$$

The expression of MSE is shown in equation (11). This is a more commonly used loss function. Where MES is the average loss function, the q^{real} is the real value of employee value, and the q^{pre} is the predicted value of employee value.

$$L = \text{MSE}\left(q^{\text{real}}, q^{\text{pre}}\right) = \frac{1}{nm}\sum_{k=1}^{N}\sum_{j=1}^{M}\left(q_{kj}^{\text{real}} - q_{kj}^{\text{pre}}\right)^2. \tag{11}$$

The training data sets, testing data sets, and predicted value can be described as equations (12)–(14). The "Train" is the mean of training data sets and the "Test" is the mean of testing data sets. The y is the predicted value.

$$\text{Train} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n), \ldots, (x_N, y_N)\}, \tag{12}$$

$$\text{Test} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m), \ldots (x_M, y_M)\}, \tag{13}$$

$$\hat{y} = \left\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_{m,} \ldots, \hat{y}_M\right\}. \tag{14}$$

4. Result Analysis and Discussion

After the BP model is established, the process of iterative training will begin. This study compares the accuracy of two deep learning predictions, the single BP neural network without LSYM layers and a BP neural network with LSTM layers. Figures 5 and 6 show the training loss and test loss under two models conditions. From Figures 5 and 6, it can be seen that the training and test losses of the BP neural network with LSTM are relatively small, which shows that the neural network with LSTM captures the temporal characteristics of enterprise performance very well.

In general, these two models have good convergence, whether it is training loss or testing loss. The two types neural network model have been adopt, which will be used to study the prediction accuracy of cost-saving rate. Meanwhile, it also could be seen from Figure 5 that the training data set and the testing data sets reach the convergence level within 500 steps, which shows that the neural network model can better fit the nonlinear relationship between corporate performance factors and the cost-saving rate very well. It can be seen from Figure 6 that the BP neural network with LSTM layer converges faster than a single BP neural network without LSYM layers, which is mainly due to the time dependence of the cost-saving rate. Moreover, the loss of a single BP neural network fluctuates in the initial stage of the testing data sets and which reaches a stable convergence level in the later stage. From the above, it can be concluded that the BP neural network can have better learning and predictive capabilities in fitting the relationship between the cost-saving rate and the enterprise performance factors. At the same time, it is necessary to fully consider the time dependence between the enterprise performance factors. The neural network will also have better learning and predicting abilities.

Figure 7 shows the predicted value and the true value of business performance of the cost-saving rate within one year. It can also be seen that the difference between the predicted value and the true value was relatively small, and the error is within an acceptable range. The predicted value of cost savings cannot only match the overall trend better with the true value but also that predict the change trend of the cost recovery rate with the number of days. For these two types of networks, the prediction error is within 5%, and the main error occurs where the cost-

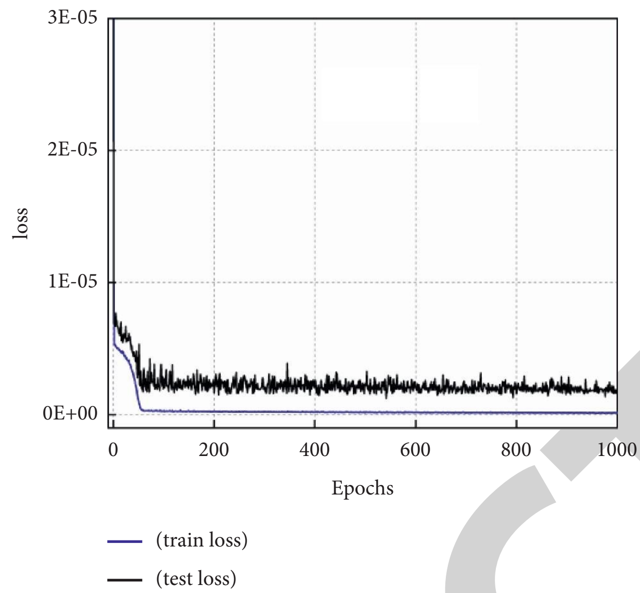


Figure 5: The training loss and testing loss of BP neural network model.

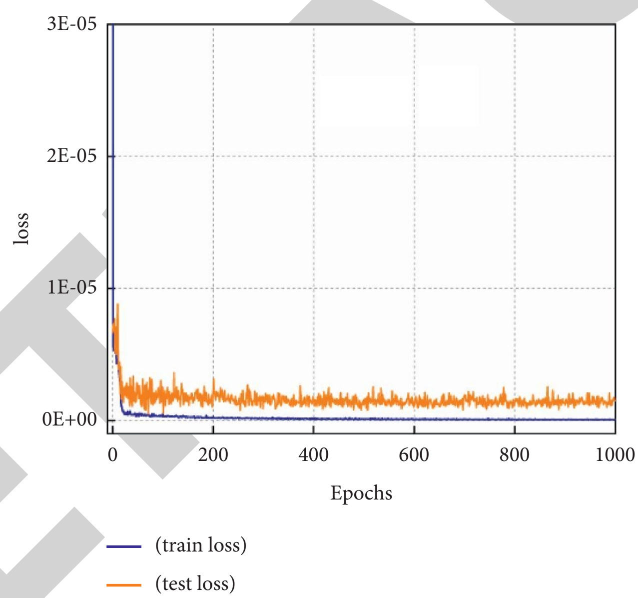


Figure 6: The training loss and testing loss of BP neural network model with LSTM layer.

saving rate changes greatly. The minimum error is only 0.8%, which is an approximately negligible error. Other errors are between 1% and 2%, and the prediction accuracy has been greatly improved. It is due to that the distribution of data set is uneven, which can be sampled in a denser place. As time goes by, the cost-saving rate fluctuates greatly, which is due to the increasing influence of enterprise performance factors, and it shows that the cost-saving rate has a clear correlation with time characteristics. It can also be seen from Figure 8 that the prediction accuracy rate of BP neural network with LSTM layers are slightly higher than the prediction performance of a single BP neural network where the cost-saving rate changes greatly.

The linear correlation coefficient curve can more intuitively reflect the fitting performance of the predicted value and the true value. The linear correlation curve reflects the distribution of the predicted value and the actual value of the enterprise performance. The closer its data value is to the $y = x$ curve, the closer the values of x and y are, which further indicates that the prediction effect is better. It can be seen from Figures 9 and 10 that the data are well distributed on both sides of the linear straight line, which shows that the predicted value is well matched with the true value. At the same time, it can also be seen that the large error of the performance prediction value is the place that deviates from the linear straight line. The value of the correlation coefficient is generally a value between 0 and 1.

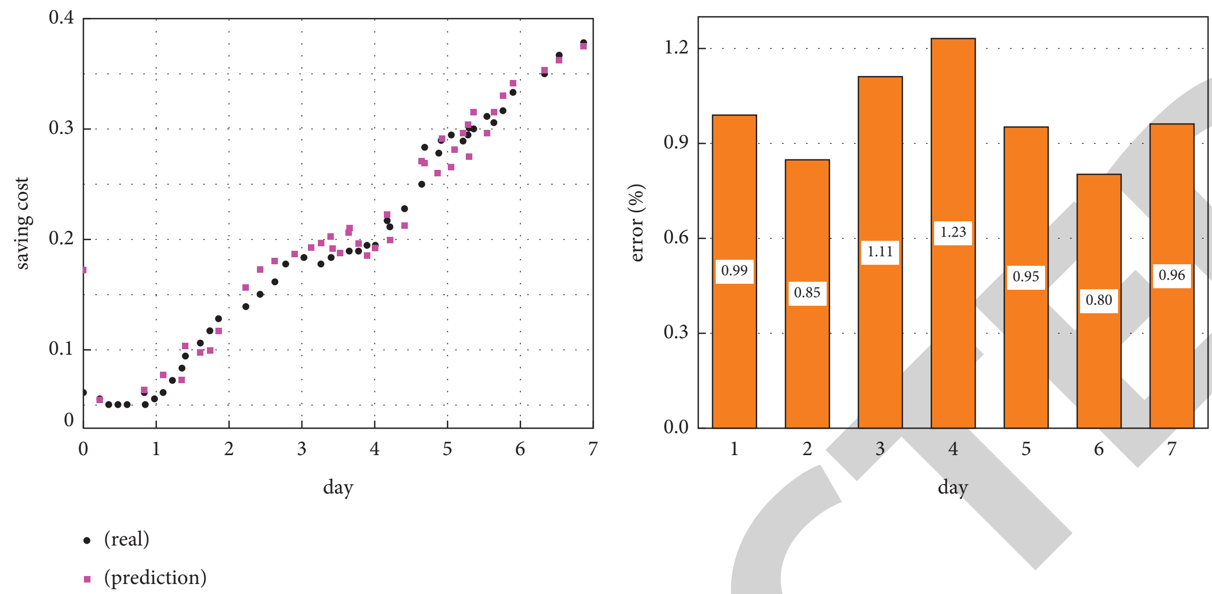


Figure 7: The predicted value and true value of BP neural network model.

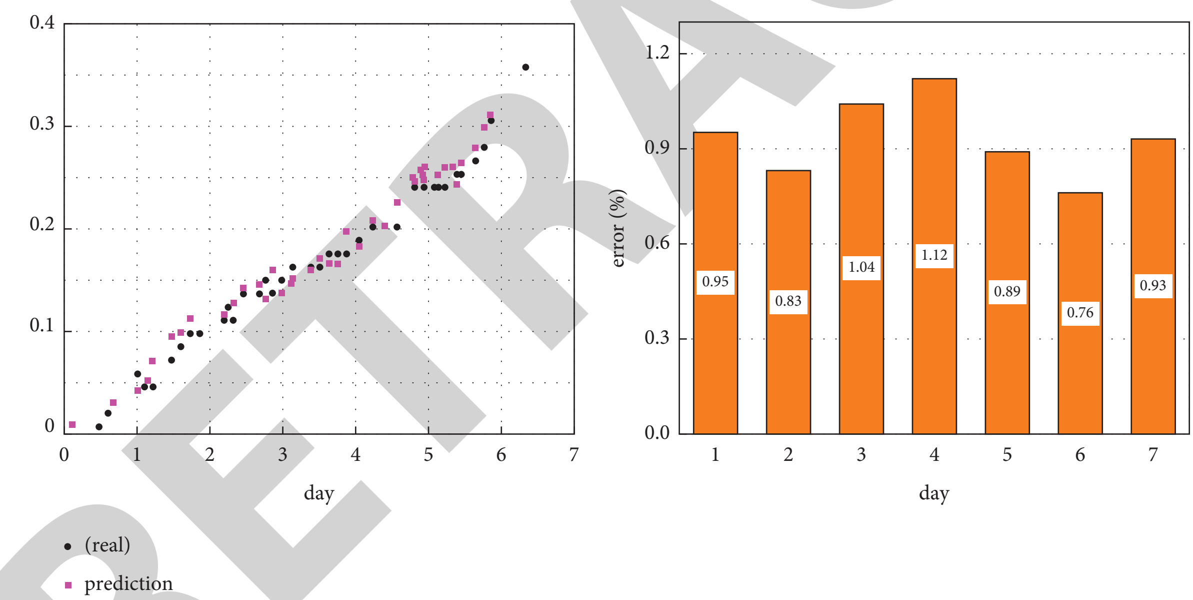


Figure 8: The predicted value and true value of BP neural network model with LSTM layer.

The closer to 1, the better the predicted value fits. Generally speaking, if the linear correlation coefficient exceeds 0.9, it means that the prediction effect is relatively stable, and if it exceeds 0.95, it means that it has better prediction performance. Most of the correlation coefficients exceed 0.9, which can indicate that the prediction performance can meet the requirements of prediction performance. The closer the predicted value on both sides of the linear fitting straight line indicates the more accurate the employee value prediction in this part. Overall, the correlation coefficients are all over 0.95, which further shows that the BP neural network has a better fitting ability to the cost-saving rate. The correlation coefficient of the BP neural network with LSTM layer exceeds 0.97, which has better accuracy than a single BP neural network. This shows that there is an obvious time correlation between enterprise performance factors and cost-saving rates. When predicting the cost recovery rate in the future, the influence of time characteristics can be fully considered, which can improve the predictive readiness and generalization ability.

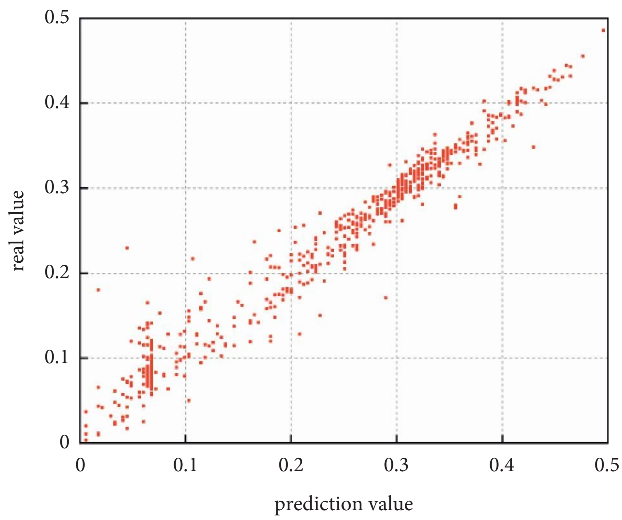


FIGURE 9: The correlation coefficients of the BP neural network model.

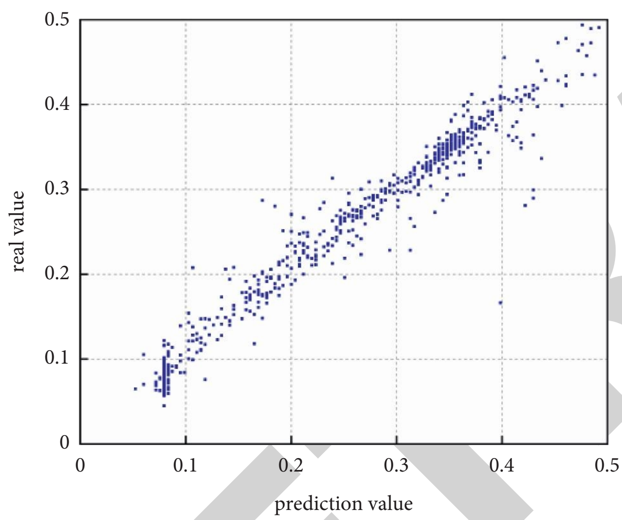


FIGURE 10: The correlation coefficients of the model with LSTM layer.

5. Conclusion

In the fourth section of this article, the accuracy and feasibility of BP neural network in predicting enterprise performance are shown in detail. Although the data set of corporate performance is highly nonlinear and the fitting relationship between them is complex, the BP neural network also predicts the future development trend of corporate performance very well, which is meaningful for corporate performance management. It is also a good model for reference.

In this research, the BP neural network has been used to predict the future trend of enterprise performance management. From the perspective of training and testing loss functions, the BP neural network model can learn the nonlinear relationship between corporate performance factors and cost-saving rates. The convergence of the loss function is faster and reaches a smaller convergence value. After considering the time characteristics, the learning ability and prediction ability of the BP neural network are improved, that is, the BP neural network has an LSTM layer. This shows that it is necessary to fully consider the influence of time characteristics when predicting the cost-saving rate through multiple factors of enterprise performance. And the learning and forecasting capabilities of the cost-saving rate will be improved. For the prediction of cost-saving rate, the BP neural network can not only match the cost-saving rate of each day well but also can better match the change trend of the cost-saving rate with the number of days. Where the cost-saving rate varies greatly with the number of days, the BP neural network with LSTM layer has better prediction accuracy than the BP neural network, which shows that there is a strong time correlation between multiple factors of corporate performance and the cost-saving rate. The overall prediction error of the cost-saving rate is within 5%, whether it is a single BP neural network or a BP neural network, this is an acceptable prediction range. This model shows good predictive ability and generalization ability. From the correlation coefficient distribution diagram, it can be seen intuitively that the BP neural network model has good predictive performance, and the correlation coefficient values are all over 0.95. The correlation coefficient R of a BP neural network with LSTM layer exceeds 0.97, which shows that this model can better fit the time characteristics between enterprise performance factors and cost-saving rates. And it can improve the predictive readiness and generalization ability. The BP neural network shows the ability to fit between the enterprise performance factors and cost-saving rate, which can provide certain reference value for enterprise performance management. Compared with BP neural network, BP neural network with LSTM layer has better performance in predicting enterprise performance characteristics, which can provide certain reference value for subsequent enterprise managers.

Data Availability

The data used in this article can be reasonably requested by readers and researchers.

Conflicts of Interest

The authors declare that there are no conflicts of interest in the study.

References

[1] J. Manyika, M. Chui, and B. Brown, *Big Data: The Next Frontier for Innovation,competition,and Productivity*, McKinsey Global Institute, 2021.

[2] E. Khalid, C. Han, A. Baker, and K. Aahed, "The relationship between purpose of performance appraisal and psychological contract: generational differences as a moderator," *International Journal of Hospitality Management*, vol. 86, no. 10, pp. 86–89, 2020.

[3] M. Flyverbom, R. Deibert, and D. Matten, "The governance of digital technology, big data, and the internet: new roles and responsibilities for business," *Business & Society*, vol. 58, no. 1, pp. 3–19, 2019.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9872735, 1 page
https://doi.org/10.1155/2023/9872735

Retraction

Retracted: The Construction of Interactive Classrooms in Colleges and Universities Based on Big Data Analysis and Benchmark Graph Neural Network

Security and Communication Networks

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] P. Shen, "The Construction of Interactive Classrooms in Colleges and Universities Based on Big Data Analysis and Benchmark Graph Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 9214022, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9214022, 13 pages
https://doi.org/10.1155/2022/9214022

Research Article

The Construction of Interactive Classrooms in Colleges and Universities Based on Big Data Analysis and Benchmark Graph Neural Network

Peipei Shen

JiuZhou Polytechnic, Xuzhou, Jiangsu 221000, China

Correspondence should be addressed to Peipei Shen; 02068@jzp.edu.cn

Received 14 February 2022; Revised 16 March 2022; Accepted 2 April 2022; Published 12 May 2022

Academic Editor: Chin-Ling Chen

In the 21st century, big data technology is bringing unprecedented changes to all walks of life. In the era of big data, new technologies, new models, and education and teaching practices are constantly looking for the best combination. Words such as interactive classroom and teaching informatization are undergoing a process of continuous enrichment of their own content with the update of new technologies and the realization of various applications. The methods of online teaching are merging with the conventional teaching practice of universities. Interactive classroom is the representative of combining the advantages of webcasting technology with classroom teaching. It uses video and audio acquisition and coding technology, C4.5 algorithm, to calculate and transmit data and analyzes information entropy-related concepts, etc., to transmit interactive classroom teaching to various spaces. This has increased students' interest in learning by nearly 50%, and teachers' evaluation scores for students have also increased to nearly twice, which shows the impact of interactive classrooms on students' self-drive. The key to the classroom does not refer to long distances, to break time and space, or to the expansion of learning groups and the unlimited sharing of educational resources. It should be the learner's self-cognition and the barrier-free interaction with the teacher, which have been reflected in the interactive classroom.

1. Introduction

Education informatization is becoming an important content of global education society. In the field of education and teaching, words such as distance teaching and teaching informatization are experiencing a process of continuous enrichment of their own content with the update of new technologies and the realization of various applications. Education and teaching under the traditional system are calling for a new supporting teaching system and innovative teaching mode that adapt to this conversion process. Teaching today not only inherits the existing characteristics of the class teaching system in the era of industrial civilization but also presents and explores new colors as the trend of educational information is approaching.

The network real-time interactive and synchronized teaching has played a powerful role in using contemporary advanced communication technology to break the geographical and time constraints of education and teaching, bridge the gap of the uneven distribution of educational resources, and narrow the gap of unfair educational development. Adjusting or reorganizing the existing school classroom teaching and learning mode has become a hot spot of current school and social education. For example, in the field of teacher education, in recent years, China's normal education colleges and universities have carried out long-distance teacher training programs through available technologies that have important strategic significance.

In the era of increasingly developed information technology, information technology has penetrated into various fields such as medical care, art design, and e-commerce industries. In the era of increasing development of information technology, universities in different regions use big data to jointly build teaching and research platforms and conduct academic exchanges between experts from various disciplines. Teachers and students of various cooperative colleges and universities collectively carry out cooperative learning and resource sharing through the online subject course learning platform, use live broadcast teaching methods for real-time interaction, and cooperate with convenient curriculum exchange online communities. These measures will effectively promote the formation of learning interest groups and research groups and expand the breadth and depth of teaching and research in universities. The development of the interactive classroom has been far-reaching along with its research. Combining the teaching mode with interaction as the core, it provides differentiated learning plans for students, creates information-based education concepts for teachers, enables teachers to successfully use information-based teaching tools, creates smooth interaction between teachers and students, and enables students to have innovative spirit be improved.

This paper first explains the research status of educational informatization and then introduces the relevant concepts and characteristics of interactive classrooms, describes the relevant concepts and formulas of big data and benchmark graph neural network technology, and puts forward suggestions to ensure the effectiveness of interactive classrooms.

Figure 1 shows the design flow chart of this paper.

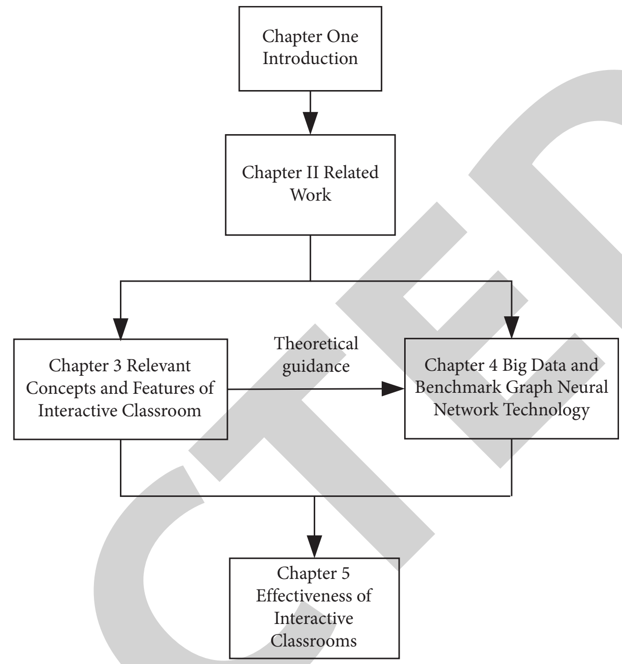


Figure 1: Design flow chart.

2. Related Work

Inequality in education in China's urban and rural areas is widespread. To solve this problem, Zhou and Xiong adopted a live broadcast class in the fifth-grade English class. Using both quantitative and qualitative methods, 90 students and three teachers from three classes in two primary schools in Yunnan Province were selected as participants. The results suggest that the use of live classrooms is feasible to solve the unfair problems in microeducation (classrooms): bridging the gap between urban and rural students in terms of scores and attitudes towards English [1]. Recent research by Rahman et al. has shown that using classroom social networking media can increase involvement and activity. However, students' attitudes towards this tool vary in terms of classroom participation, the academic performance, and overall instructional outcomes. The purpose of his research is to determine the effectiveness of real-time feeds for student classroom participation. At the completion of the lesson, students were approached to take a postsurvey to assess their impressions of class attendance after they had used live streaming in class. 50 out of 62 students (81%) responded to the survey, 70% reported that the live feed app facilitated collaborative learning and discussing, and 68% said it had increased their understanding of the topic. 66% reported that the live streaming app created a protective and anonymous environment, 64% felt it encouraged them to ask questions, and 60% of students said that the live feed did not increase their engagement [2]. Students are the masters of learning, and this should be guaranteed throughout the classroom. Therefore, for this research, the investigation of students is conducive to the protection of the status of student masters. Altinpulluk studied massive open online course (MOOC), which is among the online application of learning in which the technologies of communication connect the planet. It brought together members of the academy from the most outstanding universities to autonomous learning for themself by individuals in remote and isolated corners of the planet. While the MOOC concept is considered to be a cost-free, open online program offered in academia by respected academics from reputable universities worldwide, it comes close to the MOOC that has become so prevalent in recent years that it has drawn interest and critically analyzed through the theoretical framework. While conducting these analyses, it uses "critical posthumanism," an umbrella concept that covers many theories [3]. Regardless of the content and form of the delivery mechanism, student participation is the key to successful teaching and learning. However, in an online learning environment, involving students is a special challenge. Unlike face-to-face courses, online courses present a unique challenge because the only social existence between teachers and students is through the Internet. Khan et al. discussed building on various pedagogical approaches, various strategies for designing online study sessions can be incorporated to foster high levels of student engagement. The role of collaborative student engagement tools in the design and delivery of online courses is also discussed as well as the role of these tools in creating an atmosphere of active student participation in learning activities and active discussion [4]. Bueno et al. is discussing the development of technical teaching content knowledge (TPACK) for math teachers (preservice and in-service) who have participated in online courses using GeoGebra to develop educational applications and puzzles. The theoretical basis is based on the TPACK frame and the development of the TPACK for Mathematics teachers by none other than Margaret Niess. In this regard, three representative categories are used to construct the following meta-texts: teaching methods, classroom environment management, and teacher professional development. Through discourse text analysis, it is possible to understand the concerns of participants when using information and communication technology to create different teaching methods. They realize that dealing with numbers requires teaching innovation, the support of new technologies, and new methods of configuring and managing digital resources in the classroom [5]. During the recent COVID-19 epidemic, an online course has become an increasingly important form of student learning. Nevertheless, online tutoring does not allow for face-to-face communication in the classroom to accurately judge students' abilities. There are a number of questions and constraints such as unidirectional assessment, omission of the evaluation process, and simple evaluation forms. Therefore, Wang and Yu's research on how to establish an online course scoring system and effectively utilize the scoring mechanism has become a pressing issue. An optimization framework for proposing a process-based evaluation of online courses, which uses deep study and cofiltration techniques based on online course scoring data and student reviews, analyzes online course scoring optimization [6]. These methods provide some references for our research, which has been recognized by the public due to its relatively long time and large sample size.

3. Related Concepts and Characteristics of Interactive Classroom

3.1. The Meaning of Interactive Classroom. Introducing "interaction" into the field of classroom mode is an interactive system that takes place in a variety of situations, has a variety of forms, and has a variety of content. Traditional teaching is mostly based on the teaching method and ignores the interaction between teachers and students. Interactive classroom is to form a new type of classroom mode that focuses on student learning and develops teachers, students, and students together as shown in Figures 2 and 3.

The traditional teaching model only emphasizes the dominant position of the teacher and is completely indoctrinating teaching. Instillation education is generally called "infusion" education. It is mainly characterized by emphasizing the inculcation of knowledge by educators to students. Indoctrination education is the product of the development of human society to a certain period of time. Its formation and development are closely related to people's different understandings of knowledge, children, and teaching under certain social conditions. Students are very passive in learning, and this kind of teaching mode makes most students weary to study.

What students can harvest in the interactive classroom are class videos, test materials, questionnaires, and so on as shown in Figure 4.

It introduces the network into classroom teaching, creates an interactive classroom teaching environment based on big data and benchmark graph neural network, and develops creative interactive teaching activities to cultivate the creative ability of college students. After they have mastered the basic creative thinking methods and skills, they are guided to use the professional knowledge and skills they

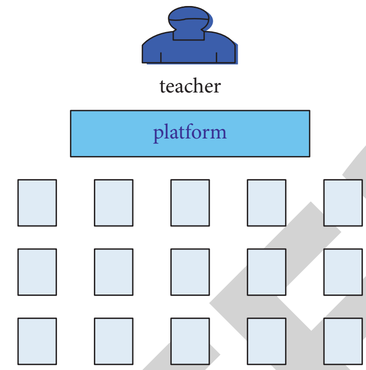


Figure 2: The teaching mode of teachers' traditional teaching.

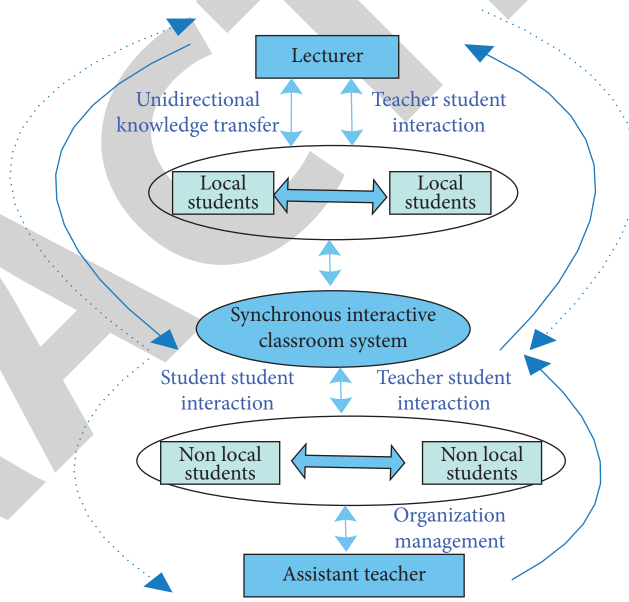


Figure 3: The lesson mode of the interactive classroom.

have learned to solve practical problems and display their creative talents [7]. The literature generated around interactive classrooms can be roughly divided into five indicator items: "same-frequency interactive classroom," "special delivery classroom," "remote synchronized classroom" or "synchronized classroom," "live classroom," and "recorded classroom." The specific data enumerating the live classroom and synchronized classroom are shown in Figure 5.

From the above data, it is not difficult to see that there is an endless stream of relevant documents generated around interactive classrooms. And in the past few years, there has been a continuous growth trend, and the number of "live classrooms" has even exceeded a thousand. In the era of big data, the research on the integration of information technology in interactive classrooms is changing from quantity to quality [8].

3.2. Status Quo of Interactive Classroom Development. Although Chinese economic development level is low and the education hardware construction is weak, the current construction of the computer network environment in

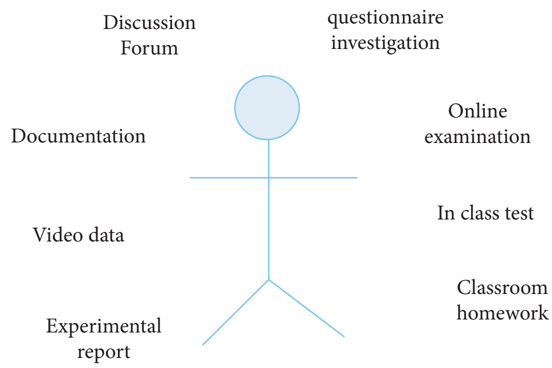


Figure 4: Student role permission system.

colleges and universities is basically complete. This provides objective hardware conditions for online classrooms [9]. The popularization rate of computer equipment in various majors of colleges and universities is selected as shown in Table 1.

From the table data, it is not difficult to find that the classrooms of information technology majors are equipped with computers and even many art fields. Majors in the liberal arts field are also equipped with computers and up to 80%. With the development of information technology, the education and teaching of colleges and universities in the era of big data are supported by computer technology. This highlights the progress of the country's science and technology and the importance of science and technology teaching.

In the process of network hardware construction, the construction of network software has gradually matured [10]. In this way, using the existing environment and resources can create a network-based interactive classroom teaching situation. And most of the teachers in colleges and universities have received educational technology training. They have basically mastered the concepts and skills of educational technology and can improve their educational technology literacy in actual teaching and scientific research. In this way, students learn in the classroom environment constructed by the network, and their innovative ability can be cultivated and their practical ability can be exercised. Based on the above objective conditions, under certain theoretical guidance, we construct an online classroom teaching environment to study the classroom behavior of teachers and students [11].

Investigating the interactive classroom in Chinese colleges and universities, there have been two teaching methods, namely, network-based and network-assisted. Taking the network as the mainstay, mainly used in distance education, and using the network as the supplement refer to the campus network teaching in colleges and universities [12]. For example, some schools have introduced the advanced People's Education Publishing House network English teaching system, which makes the integration of multimedia technology and middle-school English courses more convenient and quick, and provides a guarantee for improving the efficiency of classroom English teaching. At the same time, the two major elements of the English subject-knowledge and content urge us to make full use of modern educational technology, maximize the advantages of online English teaching, and promote the integration of information technology and curriculum.

3.3. Features of Interactive Classroom. Through the understanding of the above meanings, in addition to the general characteristics of the traditional classroom model: purpose, planning, process, etc., the interactive classroom model also has its own characteristics.

(1) It mainly adopts questioning style. Questioning in the interactive classroom model is a major and necessary form of interaction. Whether it is a problem raised by a teacher or a problem raised by a student, it will eventually be solved because the interactive classroom is the process of constantly raising and solving problems around the teaching content.

(2) Equality. According to the above understanding of the meaning of the interactive classroom model, both teachers and students are interactive subjects, and both regard each other as equal subjects [13]. Through these interactive exchanges between teachers and students, the collision of ideas and the exchange of spiritual world can be realized. This means mutual recognition, mutual equality, mutual understanding, and respect between the parties involved. Although teachers have the advantages of profound professional knowledge and ability, there should be equality between teachers and students. In traditional classroom teaching, the students' classroom behavior is required to follow the rules and obey orders. That is, the traditional thought of "the dignity of the teacher," the teacher's authority status cannot be shaken, and the teacher's right to lead the classroom cannot be overridden. As long as a student completely obeys the teacher in thought or behavior, he is a "good student" in his mind, and the teacher's words and deeds often show strictness. However, students dare not or will not ask questions to teachers. Teachers ignore the students' potential and autonomy and have no chance to express their opinions. The authority of the teacher is absolute, and the students only accept it passively. It can be seen that the teacher-student relationship under the traditional classroom model is a relationship of obedience and order, but this idea is abandoned in the historical interactive classroom.

(3) Effectiveness. Compared with the traditional classroom model, to what extent does the interactive classroom model allow students to become the main body of learning in the classroom. Under the premise of avoiding waste of time and energy to the greatest extent, the purpose of implementing interaction is to effectively complete teaching tasks while ensuring teaching quality and achieving teaching goals. It enables students to obtain in-depth development in cognition and skills, emotional attitudes, and values, that is, actual effect [14].

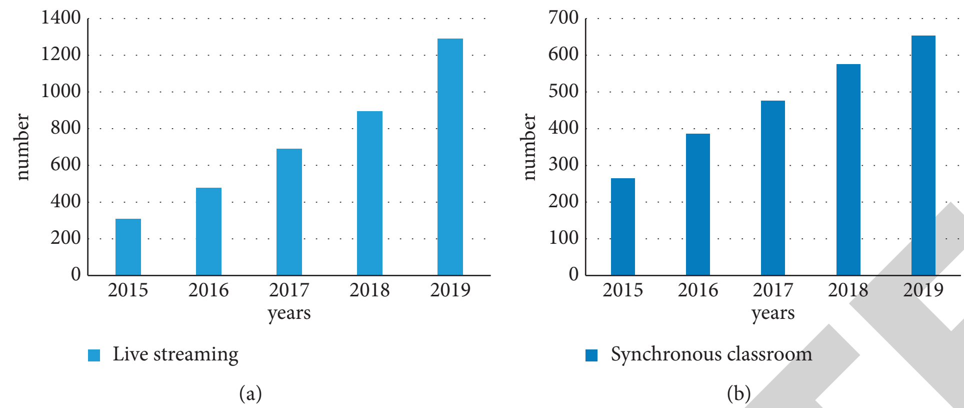


Figure 5: Specific data of live classroom and synchronized classroom.

Table 1: Computer equipment penetration rate of various majors in colleges and universities.

	University 1 (%)	University 2 (%)	University 3 (%)	University 4 (%)
Information technology	100	100	100	100
Art design	89.9	78.6	87.6	79.7
Chinese language and literature	68.9	67.9	78.9	72.6
News media	68.7	78.4	89.5	78.5
Radio host	78.9	67.9	87.5	87.2

In addition to the above, there are a lot of features that are very beneficial to teacher-student communication in the classroom, the most important and most basic is its interactivity. The characteristics of the interactive classroom mode are shown in Figure 6.

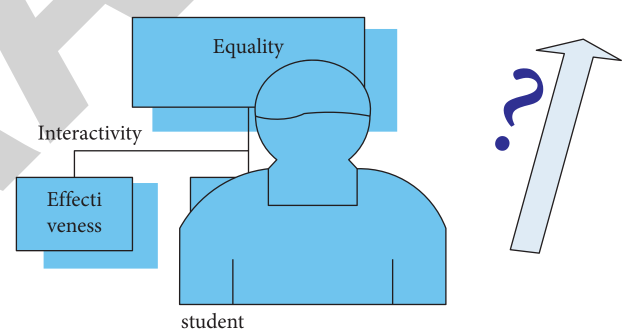


Figure 6: Features of interactive classroom mode.

4. Big Data and Benchmark Graph Neural Network Technology

4.1. Interactive Classroom Video and Audio Acquisition and Coding Technology. Video acquisition mainly refers to the conversion of live teaching activities in colleges and universities into data that can be transmitted on the network or in satellite channels through video acquisition equipment and digital signal encoding systems [15]. It uses intern recording and video guide recording and broadcasting equipment and signal lines or uses an automatic recording and broadcasting platform to form live teaching data from multimedia resources such as audio and video and teaching courseware at the teacher's teaching site. After being encoded by the streaming media server, it is transmitted to the live cloud server in real time and distributed to the remote synchronized classrooms of other colleges and universities in time as shown in Figure 7.

The video capture and encoding server of the interactive classroom is a real-time audio and video processing equipment that supports all services. It provides live video function as the core, while assisting other services such as time shifting and watching back [16]. The system integrates information source collection, coding system, stream fragmentation processing, and output format processing modules. It can encode multiple formats of audio and video signal input, output multiple streams to implement services such as live broadcasts, and rely on lower bit streams to obtain higher-quality audio and video. It is suitable for the video specifications of multiple platforms such as mobile phones and the Internet and can automatically select the appropriate output method according to the actual application scenario. When users use it, they only need to connect the network interface, audio and video interface, input and output channels, and other corresponding cables to construct various live streams. During the live broadcast, the live content can be encoded and saved as a local file. As the content of the review, it is supported in the live broadcast at the same time. Like other programs, the review supports fragmentation and streaming broadcast and supports video review at any time within 7 days [17] as shown in Figure 8.

4.2. Regression Analysis. Regression analysis is used to determine the quantitative relationship between two or more variables. And using regression analysis can find a certain rule from some actual data, such as establishing a deterministic or nondeterministic relationship between observable factor variables and dependent variables:

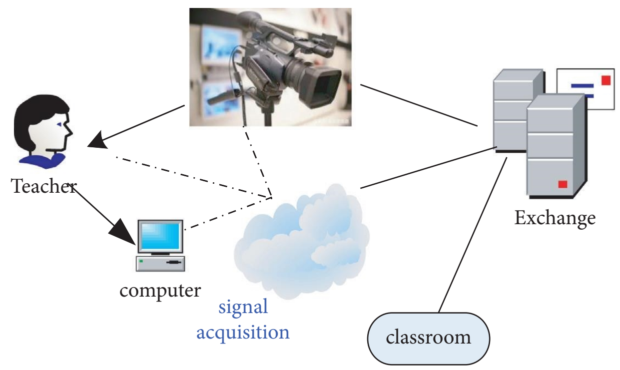


Figure 7: On-site video acquisition system for interactive classroom teaching.

$$Y = f(x_1, x_2, \ldots x_n) + \beta,$$
$$y = E(y) = f(x_1, x_2, \ldots x_n). \quad (1)$$

where β and Y are random variables, namely:

$$Y = f(x_1, x_2, \ldots x_n) + \beta = \mu_0 + \sum_{i=1}^{m} \mu_i x_i + \beta. \quad (2)$$

It can be obtained from this that, due to the uncertainty of accounting, linear regression analysis is to study the uncertainty relationship formula through a formula with a definite relationship. It needs to obtain the estimator of y by finding the estimator of μ and μ_i. At the same time, the regression data are analyzed to calculate the number of viewers in the interactive classroom on the spot, the number of interactions, and the number of online users at the same time period, etc. The calculated data are shown in Table 2.

Reflected from the data, all real-time classroom data can be presented through calculation so as to track the class information in real time.

4.3. Principal Component Analysis. The purpose of principal component analysis is to transform multiple index variables that may have correlations into a few independent linear and uncorrelated comprehensive index variables through orthogonal transformation.

Principal component analysis was first introduced by Karl Pearson for nonrandom variables, and then, H. Hotelling extended this method to the case of random vectors. The size of the information is usually measured by the sum of squared deviations or variance. The mathematical model of principal component analysis is as follows: let X be the initial variable, and the variable B needs to be required. After n observations of X, the observation data matrix can be obtained.

$$X = \begin{bmatrix} x_{11} x_{12} \ldots x_{1q} \\ x_{21} x_{22} \ldots x_{2q} \\ \ldots\ldots \\ x_{n1} x_{n2} \ldots x_{nq} \end{bmatrix} = (x_1, x_2 \ldots x_n), \quad (3)$$

in

$$x_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \ldots \\ x_{ni} \end{bmatrix} \quad (i = 1, 2, \ldots n). \quad (4)$$

Using the q vectors of the data matrix x, we get

$$B_1 = a_{11}X_1 + a_{21}X_2 + \cdots a_{q1}X_q,$$
$$B_2 = a_{12}X_1 + a_{22}X_2 + \cdots a_{q2}X_q, \quad (5)$$
$$B_3 = a_{13}X_1 + a_{23}X_2 + \cdots a_{q3}X_q,$$

$$B_q = a_{1q}X_1 + a_{2q}X_2 + \cdots a_{qq}X_q, \quad (6)$$

It is abbreviated as

$$B_i = a_{1i}X_1 + a_{2i}X_2 + \ldots . a_{qi}X_q, i = 1, 2, \ldots q, \quad (7)$$

When X is an n-dimensional vector, B is also an n-dimensional vector. In this way, the polling principle is shown in Figure 9.

It can monitor the transmission speed of live broadcast under different devices in the interactive classroom, as shown in Table 3.

The data found that the transfer speed of FTP is more appropriate and stable [18]. It created an improved version of polling-long polling, as shown in Figure 10.

The long polling principle not only enables real-time communication but also releases the network bandwidth burdened by information interaction, reducing the waste of resources [19].

4.4. Concepts Related to Information Entropy. Self-information amount: the self-information amount of information symbol a_m is defined as $I(a_m) = -\log_2 q(a_i)$. It represents the uncertainty that the recipient sends to source a_m before receiving a_m, where $q(a_i)$ is the probability of taking a value of a_m [20]. The amount of self-information reflects the uncertainty of acceptance a_m. The greater the amount of self-information, the greater the uncertainty. Information entropy: the amount of self-information reflects the uncertainty of the symbol, and the overall uncertainty of the entire source x is measured by information entropy.

$$H(X) = [-q(a_i)\log_2 q(a_1)] + \cdots [-q(a_n)\log_2 q(a_n)],$$
$$H(X) = -\sum_{i=1}^{n} q(a_i)\log_2 q(a_i). \quad (8)$$

Conditional entropy: conditional entropy H is used to measure the uncertainty of the random variable X after the receiver receives the information Y when the source X and the random variable Y are not independent of each other. The information source symbol corresponding to X is a, and the information source symbol corresponding to Y is b, then

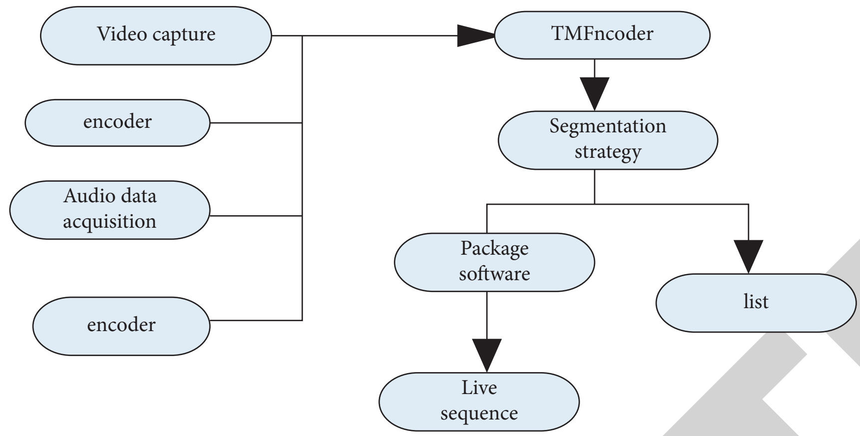


Figure 8: The internal coding process of the interactive classroom teaching acquisition server.

Table 2: The cumulative number of viewers before 16:05 in class and the number of interactive and online data in these 5 minutes.

	16:01	16:02	16:03	16:04	16:05
Number of interactions	5678	5545	4323	5443	7654
Number of people online	434	344	565	655	875
Number of spectators	890	1098	878	1298	1876

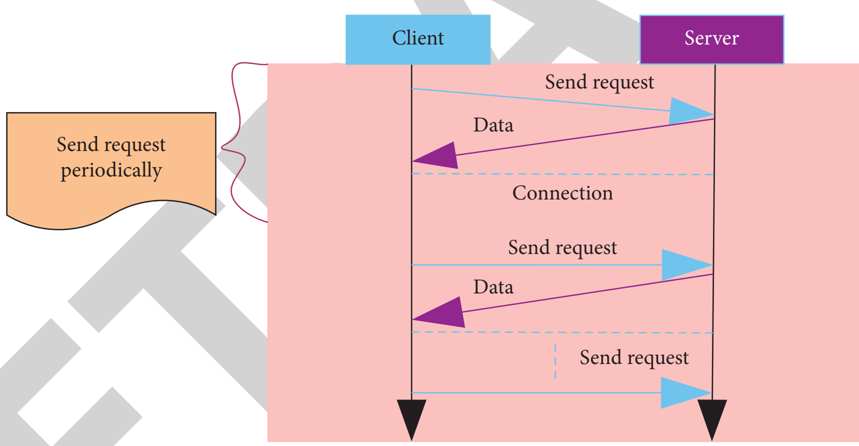


Figure 9: Polling principle.

$$\begin{aligned}
H(X|Y) &= \sum_{j=1}^{s} q(b_j) H(X|b_j),\\
H(X|Y) &= \sum_{j=1}^{s} q(b_j)\left[-\sum_{i=1}^{n} q(a_i|b_j)\log_2 q(a_i|b_j)\right],\\
H(X|Y) &= \sum_{j=1}^{s} q(b_j) q(a_i|b_j)\log_2 q(a_i|b_j),\\
&\quad -\sum_{j=1}^{s}\sum_{i=1}^{n} q(a_i, b_j)\log_2 q(a_i|b_j).
\end{aligned} \tag{9}$$

Average mutual information amount: it represents the amount of information about X that signal Y can provide, namely

Table 3: Speed monitoring of different transmission servers.

	12:01 (bps)	12:02 (bps)	12:03 (bps)	12:04 (bps)
Dell	12	14	15	16
TP-LINK	13	34	13	23
FTP	56	45	49	56

$$I(X|Y) = H(X) - H(X|Y). \tag{10}$$

4.5. C4.5 Algorithm Calculates the Transmission Data. The fundamental idea of the C4.5 method differs from the ID3 method in that the C4.5 technique selects the test attributes with the greatest possible information gained rate. In the

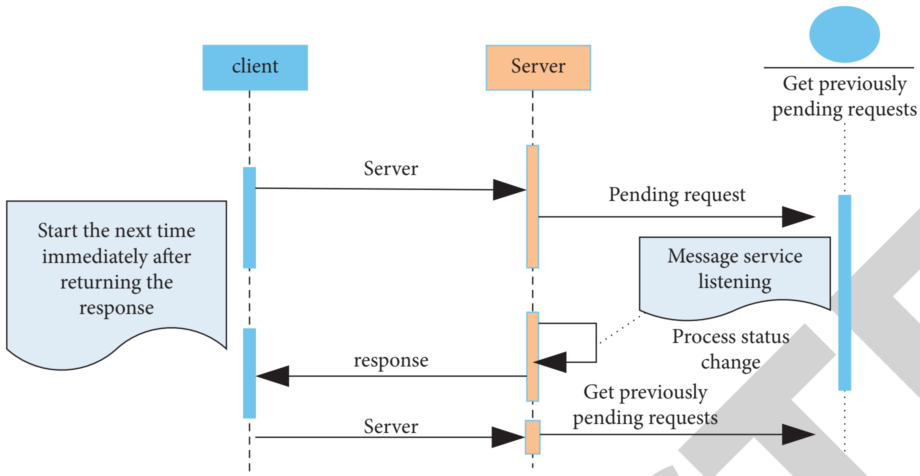


Figure 10: Long polling mode.

sample set T, if the parameter a has n properties, then the property values are recorded as $a_1, a_2, \ldots, a_n$, and the number of samples corresponding to the value of a is n_i, respectively, and

$$n_1 + n_2 +_3, \ldots, n_k = n, \tag{11}$$

where n is the total number of samples. The entropy value H of attribute a is the price that needs to be paid to obtain the information about attribute a of the sample, namely

$$H(X, A) = -\sum_{i=1}^{k} q(a_i)\log_2 q(a_i) \approx -\sum_{i=1}^{k} \frac{n_i}{n}\log\frac{n_i}{n}. \tag{12}$$

The information gain rate is defined as the ratio of the average mutual information to the cost of obtaining information, namely

$$E(X, a) = \frac{I(X, a)}{H(X, a)}. \tag{13}$$

This kind of transmission is actually "two flowers bloom, one branch on each table." Each data stream of remote live classroom teaching takes a closed line, each remote receiving school student feedback information data real-time transmission uses another line, and the two form a synchronous teaching interactive classroom transmission system through the consistency and continuity of time as shown in Figure 11.

4.6. University Networking Synchronous Live Teaching System. In the specific teaching live broadcast, the live class teaching situation captures the live teachers and students' pictures and audio through the hypothetical camera position and sound collection equipment. The teacher's computer courseware desktop signal is connected to the director station through the signal line. In the live teaching site, the teaching assistant uses the camera to record the whole course of the class and record the audio and video information on the spot. At the same time, multiple formats of teaching materials can be connected to the system through the director cut control and then transmitted to the streaming media server through the transmission signal line. After the live video and audio are converted by a streaming media encoder, they are packaged into a signal and transmitted to the cloud server of China Education Television Station through the network. It distributes the collected data streams live through cloud service groups with nodes all over the country and finally integrates a channel of live teaching data to output to the classrooms of different colleges and universities in various places. Distance learners concentrate on listening to the lectures and can communicate with the teachers and students on the live broadcast site through assistants to make up for the lack of classroom interaction. Learners can also log on to the Netlink public service platform to conduct in-depth discussions about courses at any time and interact with teachers on the platform as shown in Figure 12.

5. The Effectiveness of Interactive Classrooms

5.1. Promoting Education Equity. By applying big data and other educational technologies to the teaching of interactive classrooms in colleges and universities, the communication between students and teachers is more convenient and efficient, sharing good teachers, sharing resources, mutual exchanges, and cooperative exploration, and the classroom is more vivid.

In terms of time, communication between teachers and students is no longer limited to after class. In the process of teacher's explanation, students can also put forward their own questions and opinions at any time, question the teacher, and diverge students' thinking. This allows excellent teacher resources to spread and flow across provinces, cities, and even across the country.

In space, with the addition of educational technology, better learning materials, and more learning partners, these advantages can to a large extent allow children in underdeveloped areas to start farther and have more possibilities for good development. This allows excellent educational resources to benefit more places, share the beautiful learning life brought about by science and technology, and jointly promote educational equity.

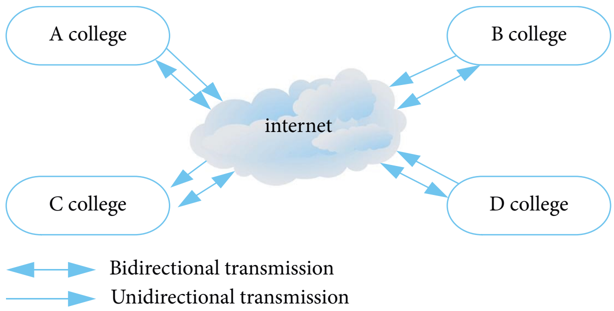


Figure 11: Networking synchronous live classroom teaching transmission system.

Combining the interactive classroom teaching concepts of the two places, the planned classroom structure of the two places is shown in Figure 13.

Home-school cooperation is the foundation of student progress, so parents must be guaranteed the right to speak. It counts the evaluation of interactive classrooms by parents from freshman to junior year in a college as shown in Table 4.

This statistic is based on the dislike, general, approved, and very approved attitudes of students' parents on the evaluation of interactive classrooms. Whether it is a freshman or a junior, parents find that their acceptance of interactive classrooms is still very high, and nearly half of them agree.

In addition to the important role of students in the classroom, the evaluation of teachers is also worth studying. In the same way, the teachers of this school make a ten-point evaluation of the students' performance before and after the interactive classroom, such as the degree of enthusiasm, the degree of interaction, the quality of homework, and the performance feedback. The data are shown in Figure 14.

According to the data, the performance of students in interactive classrooms, the degree of enthusiasm, the degree of interaction, the quality of homework, and the score feedback have increased by nearly 50% compared with the previous ones. Good class performance can have high-quality class effects and produce excellent academic results. Interactive classrooms promote a virtuous circle of student learning.

5.2. Increasing Student Interest. People often use words like "like" and "love" to express a person's identification with a certain person, a certain thing, or a certain activity. For example, children like to observe ants moving, girls like to play piano, and boys like cars. People like this tendency to explore things spontaneously and enthusiastically for certain activities from the heart, and people call it interest. Only when students have a strong interest in learning a course, will they be eager to learn more about it. When he encounters difficulties and setbacks in the learning process, he will not feel irritable and helpless, but will actively think, actively communicate with others, engage in ideological collisions, break through difficult problems, and improve himself. The time of each class is usually 40–45 minutes, and the spirit of the students cannot be focused on learning all the time. When the concentration is not concentrated, if the learning content is obscure, it will inevitably cause the understanding and memory of the abstract content to be imperfect. Modern educational technology can improve this problem because of its vivid and flexible characteristics. It can use pictures, videos, audios, and animations to make boring single text or complex images more vivid and intuitive, arouse students' interest in learning, stimulate internal motivation, and make students' learning fun. For example, in the actual middle-school classroom English teaching, the teacher can first listen to some English songs related to the new class to the students and guide the students to sing along, or they can ask the students who can sing to lead everyone to learn to sing English songs. It starts new learning in a cheerful and harmonious classroom atmosphere, allowing students to accept English learning naturally. It then guides students to learn specific vocabulary or sentences. However, the creation of English teaching situation is not limited to the introduction link. Teachers can also use educational technology in the follow-up consolidation, practice, and even summary, to create a vivid but not messy English teaching situation to improve the learning effect of students. The statistics of freshman to sophomore students' interest in interactive classrooms are shown in Figure 15.

5.3. Improving Classroom Efficiency. In the era of big data, personalized interactive courses integrate information application tools such as personalized learning and smart classrooms to efficiently provide teachers with scientific, reasonable, and personalized teaching design. An efficient and vivid fusion class is to combine the typical mistakes that are easy to make, focus on the common problems existing in most students, and enrich the teaching content. Teachers and students communicate in multiple directions to keep abreast of the students' learning situation to ensure that they learn how to teach and grasp the key points of teaching flexibly. The standardization of interactive classrooms is to realize the integration of teachers from preclass preparation, in-class teaching, after-class assessment, and guidance. This makes classroom interaction more efficient, evaluation feedback is more timely, and teaching goals are clearer. Interactive classroom has the advantages of rich teaching content, personalized learning, real-time performance analysis, diversified teaching methods, support for review at any time, and growth records. This makes the classroom capacity increase significantly, and the teaching efficiency is significantly improved. These characteristics show that high-quality and efficient teaching is high in density and fast-paced, and such new classrooms have high requirements for teachers and students. With the development of information technology, the connotation of curriculum resources has also expanded. For example, the more commonly used teaching methods in primary school Chinese classrooms include lecture method, problem-inquiry method, situational teaching method, and reading instruction method. Figure 16 shows the usage frequency of teachers' teaching methods before and after the interactive classroom.

After teachers have accepted interactive classroom teaching, the use of situational teaching methods and

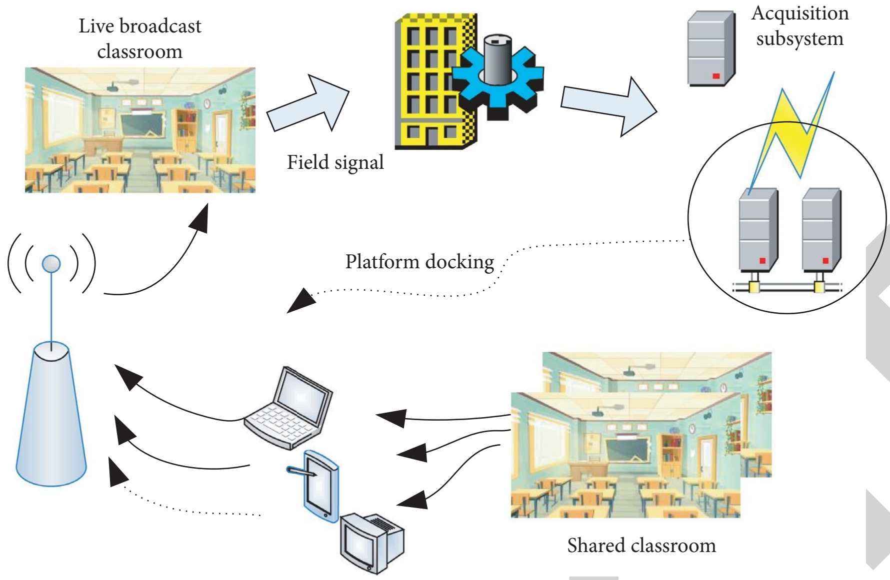


Figure 12: Architecture diagram of interactive classroom teaching system in colleges and universities.

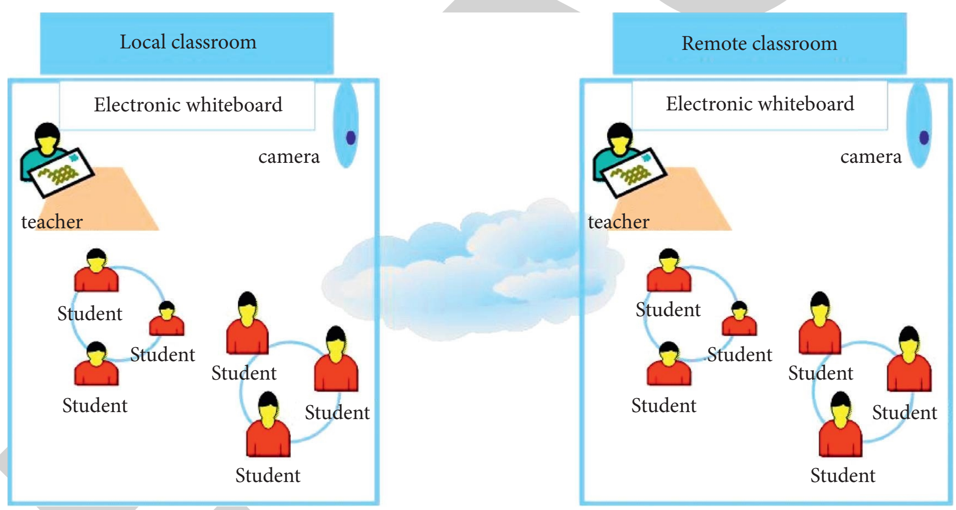


Figure 13: Floor structure diagram of interactive classroom.

Table 4: Evaluation of interactive courses by freshman to junior students in a university.

	Dislike (%)	Commonly (%)	Approval (%)	Very recognized (%)
Freshman	5.7	12	36.4	45.9
Sophomore	6.7	15	36.8	41.5
Junior	4.3	17	33.7	45

problem-inquiry methods has increased, which has greatly overturned the previous situation of mainly teaching methods.

5.4. Realizing Teaching Students in accordance with Their Aptitude. In traditional classroom teaching, because there may be gaps between good students and poor students in many aspects, it is difficult to achieve a balance between the two. This causes teachers to be helpless when facing students of different learning levels and unable to take care of both good students and poor students at the same time. This makes both sides unable to achieve the same goal, which is also a problem that has plagued many teachers for many years. Until the emergence of educational technology, it is possible to solve this difficult problem. After applying computer technology in teaching, teachers can establish a systematic autonomous learning environment for students, which is a multilevel and multichannel environment. The

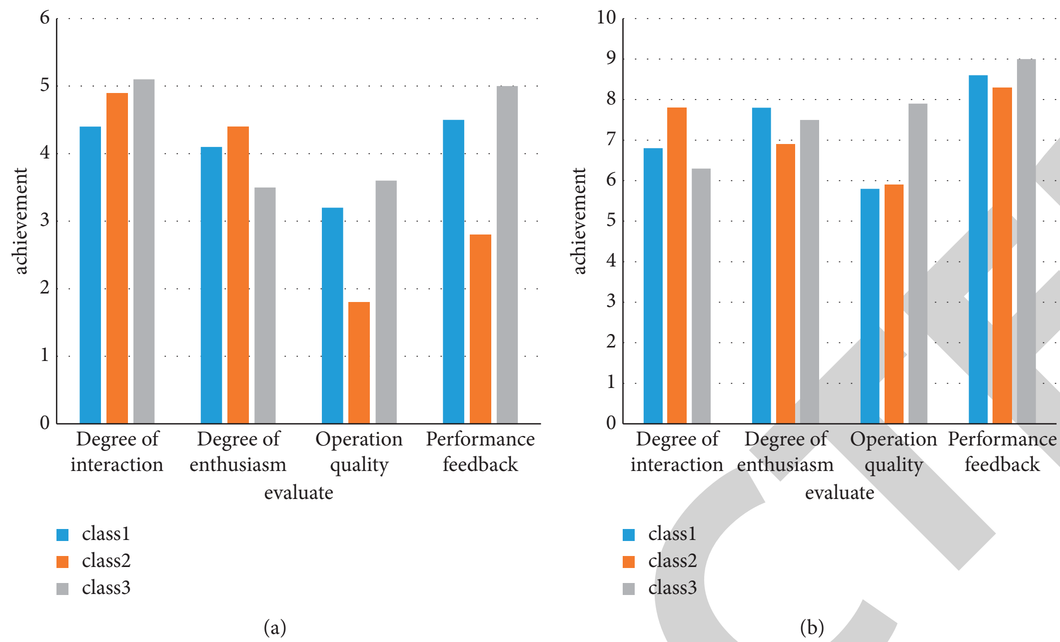


Figure 14: Teachers' evaluation of students' performance before and after the interactive classroom.

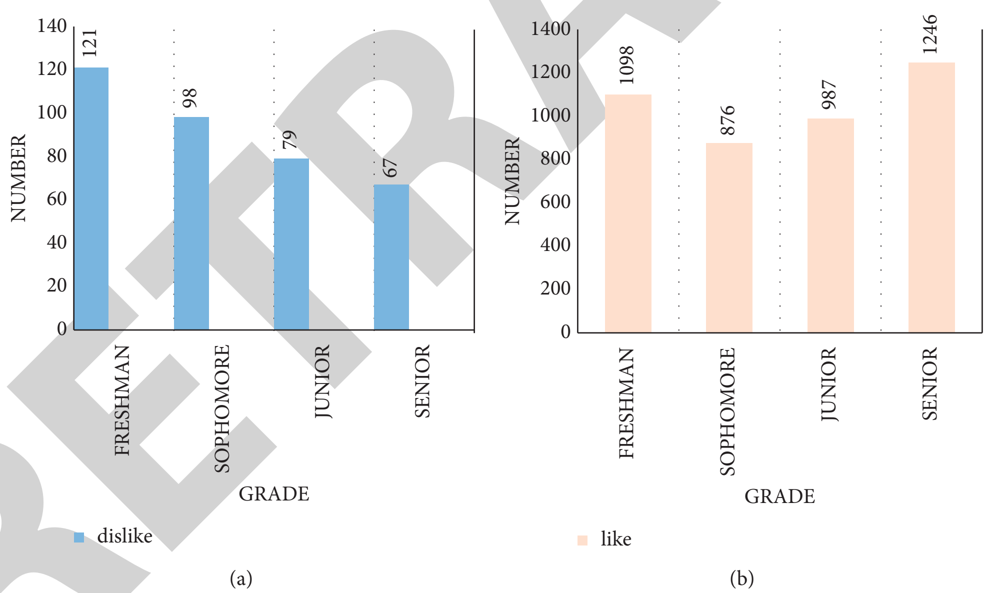
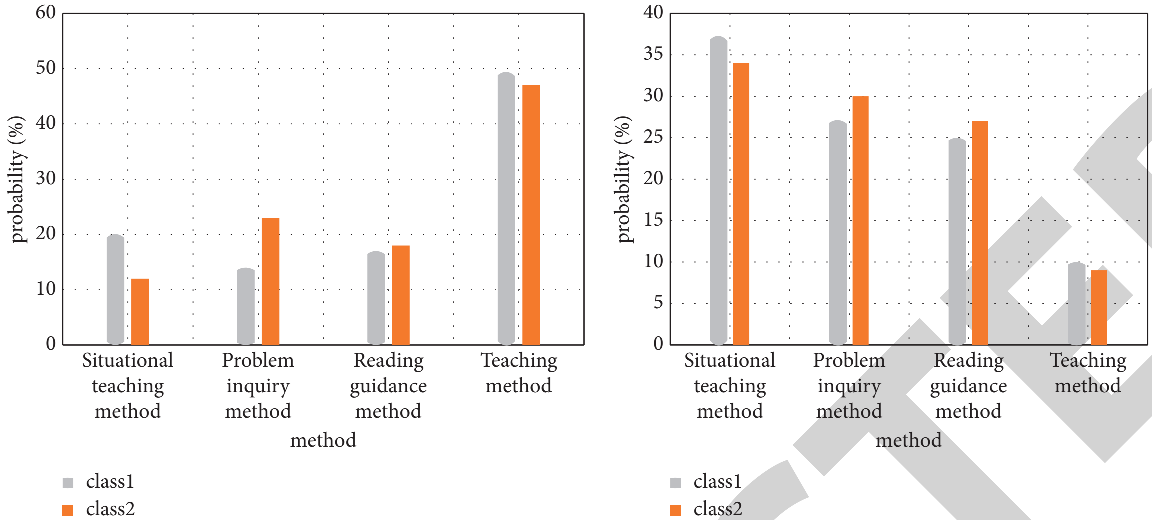


Figure 15: Freshman to sophomore students' interest in interactive classrooms.

exercises are divided into several levels from low level to high level, and students choose according to their actual situation. Students have fully mastered the lower level and then enter the higher level, thus gradually progress and lay a solid foundation. Of course, for very good students, we can also use the rich network resources to deepen the difficulty.



(a) (b)

Figure 16: Teachers' teaching methods before and after the interactive classroom.

6. Conclusion

Economic development and the continuous emergence of new technologies have promoted the process of education informatization. Educational informatization makes it possible to establish, share, and develop together regional resources. Quality education is the main theme of the current classroom teaching reform. The quality of classroom teaching is to face all students; develop the overall quality of students; let students actively discover, research, and analyze problems; solve problems; and take the initiative to construct the meaning of knowledge. Interactive classroom is the product of the combination of big data information technology and education and teaching. The teaching system centered on "interaction" strengthens the connection between students, teachers, and academic staff through network information technology. This has played an indispensable role in the field of school education. Through the integration of teaching system, teaching content, teachers, and students, a harmonious teaching environment and a good learning atmosphere are created. This enables students from different regions and schools to cooperate in discussing and solving problems, which is conducive to inspiring students' thinking and is of great help to the improvement of their learning efficiency. It has the prospect of scientific, efficient, and sustainable development. Introducing the network into classroom teaching can create an interactive classroom teaching environment based on big data and benchmark graph neural network, carry out creative interactive teaching activities, and cultivate college students' innovative ability. The interaction model between teachers and students needs to be enriched and perfected by long-term practice, and subsequent researchers should design more interactive models that are more conducive to promoting the interaction between teachers and students according to the needs of teaching activities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Y. Zhou and Y. Xiong, "Live broadcast classroom," *International Journal of Distance Education Technologies*, vol. 15, no. 3, pp. 31–46, 2017.

[2] A. Rahman, Y. Sahakian, S. Kane, M. Abel, and R. Deshmukh, "Live-feed use in the classroom as an engagement tool," *Currents in Pharmacy Teaching and Learning*, vol. 11, no. 12, pp. 1303–1308, 2019.

[3] H. Altinpulluk, "REVIEW: posthumanism and massive open online course contaminating the subject of global education," *The Turkish Online Journal of Distance Education*, vol. 19, no. 1, pp. 247–249, 2018.

[4] A. Khan, O. Egbue, B. Palkie, and J. Madden, "Active learning: engaging students to maximize learning in an online course," *Electronic Journal of E-Learning*, vol. 15, no. 2, pp. 107–115, 2017.

[5] R. W. d. S. Bueno, D. Lieban, and C. C. Ballejo, "Mathematics teachers' TPACK development based on an online course with geogebra," *Open Education Studies*, vol. 3, no. 1, pp. 110–119, 2021.

[6] Z. Wang and N. Yu, "Education data-driven online course optimization mechanism for college student," *Mobile Information Systems*, vol. 2021, no. 1, Article ID 5545621, 8 pages, 2021.

[7] K. Huang, X. Ge, and V. Law, "Deep and surface processing of instructor's feedback in an online course," *Educational Technology & Society*, vol. 20, no. 4, pp. 247–260, 2017.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9859568, 1 page
https://doi.org/10.1155/2023/9859568

Retraction

Retracted: Research on Precision Teaching Model of Ideology Course Based on Collaborative Filtering Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Li, "Research on Precision Teaching Model of Ideology Course Based on Collaborative Filtering Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5272779, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5272779, 10 pages
https://doi.org/10.1155/2022/5272779

Research Article

Research on Precision Teaching Model of Ideology Course Based on Collaborative Filtering Algorithm

Jinshan Li

Xingjian College, Xijing University, Xi'an, Shaanxi 710123, China

Correspondence should be addressed to Jinshan Li; 20020045@xijing.edu.cn

Received 18 February 2022; Revised 11 March 2022; Accepted 21 March 2022; Published 12 May 2022

Academic Editor: Chin-Ling Chen

The ideological and political course should not only keep the academic rationality and political nature of the course itself, but also take into account the characteristics of colleges and universities and students' growth and development needs. At present, there are some problems in the curriculum of ideology, such as mechanical rigidity, weak pertinence, lack of synergy, and inability to form a personalized collaborative and precise education mechanism. Aiming at related problems, this article constructs an accurate teaching model of ideological and political course based on collaborative filtering algorithm. First, the public test set of recommended fields is used to test and verify the effectiveness and practicability of the algorithm. For the data sparseness and cold start of collaborative filtering algorithm, the course feature attributes and attribute value preference matrix are used to solve the problem, and the similarity is calculated offline, so as to realize the real-time recommendation and accurate teaching of the course. In order to verify the effectiveness of this method for precise teaching, we conducted a test. The test results show that the precise teaching model has a positive effect on the improvement in students' academic performance. The method proposed in this article realizes the identity transformation of students from passive acceptance to active construction, and the teaching effectiveness of ideological and political course is effectively improved.

1. Introduction

Due to the continuous progress of information technology (IT), people pay more and more attention to educational informatization [1]. In the development process of educational informatization, the realization of precise teaching is the urgent need of educators and educatees. Accurate teaching can significantly improve learners' learning efficiency [2]. The advantage of traditional education is that teachers can know students' understanding of knowledge through students' reactions in the classroom teaching process and decide the key points of students' guidance [3]. As a public compulsory course, politics is taught in large classes in most schools at present, which easily lead to teachers' low attention to students in class and cannot meet the individualized development needs of students. And classroom teaching is old-fashioned, and students' acceptance is not high, so it is difficult to achieve accurate teaching [4]. Its disadvantages are mainly reflected in five aspects: ① the teaching mode lacks intelligence, and the courses are piled up in a single way; ② the lack of communication and information feedback mechanism; ③ the lack of effective teaching guidance mechanism; ④ the learning process lacks necessary monitoring and the teaching evaluation function is not strong; and ⑤ the lack of effective and perfect theoretical guidance. Therefore, it is difficult to achieve the ideal teaching effect. Through the analysis of the traditional teaching process, we can transplant its advantages into the network teaching process to realize intelligent teaching, that is, we can learn from the traditional teaching idea of teaching students in accordance with their aptitude [5]. Personalized and hierarchical teaching is an important guarantee to improve the teaching quality and classroom effectiveness of ideology courses. The development of IT provides technical support for collecting, storing, and analyzing students' learning data.

As the main channel and front of ideology work in universities, the ideology education course is an important way to practice the education mechanism in universities and also an important part of the theoretical course system in universities [6]. It runs through the whole process of higher education and teaching and is an important institutional carrier for universities to train socialist builders and successors. With the in-depth application of network technology in the field of education and the continuous improvement in learners' requirements, the stereotyped teaching mode and fixed learning interface cannot reflect the teaching idea of teaching students in accordance with their aptitude [7]. Generally speaking, it is the lack of intelligent thought in the teaching process. In order to solve this problem, the intelligent network teaching system came into being. For school students, some students have poor self-discipline, and it is difficult for ideology teaching to achieve the desired teaching effect. It is particularly important to enhance the attractiveness of classroom teaching. The use of IT provides an opportunity to improve the boring classroom atmosphere [8]. With the support of IT, teachers of ideology courses can introduce personalized teaching resources and students can learn online. The system will automatically record students' learning on track, and learning behaviors such as daily learning progress, learning duration, and mastery level exist in the form of data [9]. Teachers can conduct all-round monitoring, and students' ideological status and learning status become visual and quantifiable. Teaching can be targeted according to different students' specific learning conditions, and teaching methods can be adjusted accordingly [10]. The traditional teaching method of ideological and political course has great limitations, and the traditional collaborative filtering algorithm has problems such as data sparseness and cold start, which restricts the further development of collaborative filtering algorithm and ideological and political teaching. This article puts forward corresponding solutions to solve the problems such as data sparseness, cold start, and real-time performance. It provides a solid foundation for the next step to realize the precise teaching of ideological and political courses individually according to students' interest preferences and learning needs.

Ideology education plays a fundamental role in strengthening the value guidance of college students, strengthening the publicity and education of patriotic dedication, and establishing the all-round education mechanism for all employees [11]. In the process of informatization of ideology education, the problems of teaching mode, teaching efficiency, and sharing of teaching resources have been well solved, but there are still some problems [12]. Collaborative filtering (CF), also known as social filtering, is mainly divided into two categories: global-based CF and model-based CF [13]. The basic idea of CF algorithm is that users with similar interests may like the same items, or users may have similar preferences for a product similar to the one they are interested in [14]. In fact, it is a typical method of using collective wisdom. In order to realize the precise teaching of ideology course, this article improves the shortcomings of traditional CF algorithm. It also strengthens the real-time performance of the recommendation model of ideology courses in colleges and expands the applicable scope of the recommendation model of ideology courses in universities. The specific chapters of this article are arranged as follows:

First, this article discusses the research background and current situation of precision teaching of ideological and political courses and introduces the related theories and key technologies of precision teaching. This article focuses on the related content of collaborative filtering algorithm, puts forward the concept and calculation formula of tag similarity between items, and weights the tag similarity and score similarity between items to generate the final similarity calculation formula. Then, with the final similarity calculation formula as the parameter, the prediction scoring formula is used to predict the user's scores for the items, and the N items with the highest scores are selected and recommended to the user. Education and teaching data are used to carry out experiments. Finally, the experiment results show that the collaborative filtering algorithm can provide more accurate prediction, and the performance of the improved algorithm is better than that of the original algorithm.

2. Related Work

Based on the research on collaborative filtering algorithms and precise teaching of ideological and political courses, Knight started from actual projects and used tags to supplement and correct the shortcomings of traditional project-based collaborative filtering algorithms [15]. In order to solve the bottleneck problem restricting the teaching reform of ideological and political courses in colleges and universities, Zhou et al. accelerated the pace of classroom teaching reform [16]. Efficient teaching methods are adopted to lead the teaching of ideological and political courses to achieve precise advancement. Yang and Xie combined "precision teaching" with the teaching of ideological and political courses in schools and practiced the concept of precision teaching from the three paths of precise teaching objectives, programmed teaching process, and digital learning evaluation [17]. Li designed an intelligent teaching recommendation system based on collaborative filtering algorithm, realized the "one-to-one" teaching mode, and verified its effectiveness through experiments [18]. Starting from the opportunities and challenges faced by the precise teaching of ideological and political courses, Yang and Li aimed to solve the contradiction between supply and demand in the teaching of ideological and political courses, improve the quality of teaching, and truly play the important role of ideological and political courses in implementing morality and cultivating people [19]. Wang analyzed the current situation of the course selection system in colleges and universities and applied related technologies such as collaborative filtering algorithm, data mining technology, and personalized recommendation technology to the system to promote the optimization of the personalized recommendation system for course selection [20]. Wu et al. designed an intelligent teaching recommendation system based on collaborative filtering algorithm according to the traditional teaching idea of teaching students in accordance with their aptitude [21]. Yang believe that through the precise teaching reform of ideological and political courses, schools can teach ideological and political courses in a targeted manner, which can effectively stimulate students' enthusiasm for learning [22].

In this paper, the related literature is deeply studied, and aiming at the disadvantages of traditional education mode, the precise teaching model of ideological and political course is constructed by the collaborative filtering algorithm. Then, based on the actual needs of the precise teaching of ideological and political courses in colleges and universities, by introducing the gradual forgetting curve based on the timeliness change of users' interests, the disadvantages of the traditional collaborative filtering algorithm, such as low efficiency, weak adaptability, and novelty rejection, are well solved. At the same time, the characteristic attributes of users and projects are obtained, and based on them, the similarities among users, projects, and users-projects are calculated to obtain similar neighbors. Finally, the missing values are predicted according to the scores of similar neighbors so as to solve the sparsity of data. Experiments show that this method is accurate and practical, and the real-time performance of the system is high.

3. Methodology

3.1. CF Algorithm. Among all personalized recommendation systems, the application effect and situation of CF system are the best. It is also the most widely used method [23]. CF algorithm uses group intelligence to make fuzzy recommendations and makes personalized and relevant recommendations based on common points of interest in the communication range. CF recommendation, as a popular information filtering technology, can filter and analyze the filtered content, so as to analyze users' interests and improve information service quality. The core idea of the technology is social filtering, that is, users are divided into groups, and then specific users are recommended what other users in their group are interested in.

CF algorithm has good adaptability, and the recommendation rules can be adjusted adaptively according to different points of interest. CF algorithm is still effective for long tail keywords, which can better solve the problem of identifying and recommending interest points in long catalogue. In the actual algorithm, the division of user interest groups comes from the user's historical score data, which is analyzed by similarity to sort out the neighbors similar to the target user, and the items of interest to the neighbors are recommended to the user [24]. CF algorithm only depends on the user's single-dimensional behavior, does not expand the dimension, does not need to know the content of recommendation and prediction deeply, and has wide practical applicability. According to different points, CF recommendation algorithms can be divided into two categories: ① CF algorithm based on users and ② CF algorithm based on project. Project-based CF algorithm is based on the idea that the original CF algorithm based on user behavior calculates the similarity between projects to find the nearest neighbor of recommended objects. It is a widely used algorithm in personalized recommendation system at present.

Project-based CF algorithm establishes the list relationship between commodities by users' personalized scores of a certain commodity and then recommends users and predicts users' interest points based on the list relationship between commodities. The core of the user-based CF algorithm is to find the nearest neighbor set of the target user and then use the score of the nearest neighbor as a reference to predict the user's interest. The algorithm has simple process and high accuracy, so it can guarantee high recommendation quality when the scoring data are relatively complete [25].

User-based CF algorithm is to recommend resources that users may like to target users by calculating the similarity of their access behaviors [26]. If the obtained user behavior data are classified information, then clustering algorithm is generally used. For numerical information, matrix decomposition algorithm is usually used. The project-based CF algorithm is taken as an example to illustrate the basic principle of traditional CF algorithm (Figure 1).

The algorithm first generates the nearest neighbor set of the target project and then predicts the user's score of the target project according to the user's score of the nearest neighbor set of the project [27]. CF algorithm also has inherent defects, mainly because it adopts the cold start mode of initial data, and a large number of user behavior lists need to be collected at the start-up stage of the algorithm, which is not suitable for some scenes lacking initial data. Because CF algorithm uses swarm intelligence to make fuzzy recommendation, it often fails to give the basis and mechanism of recommendation, and it is not suitable for some logical applications. User-based CF algorithm is only suitable for systems with few users. If the number of users is very large, the cost of calculating the similarity matrix of users' interests will become very high. The increase of the space complexity and time complexity of its operation and the increase of the number of users are similar to the square relationship. The recommendation results of this algorithm will not be updated immediately with the new behaviors of users, and it is difficult to provide convincing recommendation explanations for users.

Corresponding to the schematic diagram of project-based CF algorithm, first, different users' scoring tables for several commodities are established to obtain the similarity between several commodities, and then Euclidean distance table is obtained by Euclidean distance evaluation. Item-based CF algorithm recommends products to users by calculating the similarity between items. Because the similarity between projects is relatively fixed and easy to calculate, this algorithm is adopted in most systems.

3.2. Accurate Teaching of Ideology Course. "Precision teaching" is put forward on the basis of Skinner's behaviorism learning theory. Accurate teaching of ideology courses requires teachers to set teaching objectives accurately, to use IT to support them, and to focus on the classroom according to the actual situation of students as well as grasp the teaching objectives and teaching content accurately and build a scientific teaching process. In order to achieve the expected teaching goal and improve the accuracy of ideology teaching, teaching accuracy should not be limited to the accuracy of teaching methods but should emphasize the all-round and accurate implementation of teaching ideas, teaching design, teaching process, and teaching effect evaluation. With the continuous development of IT, the implementation of precision teaching has technical support. Accurate teaching is conducive to the mutual penetration of various elements in

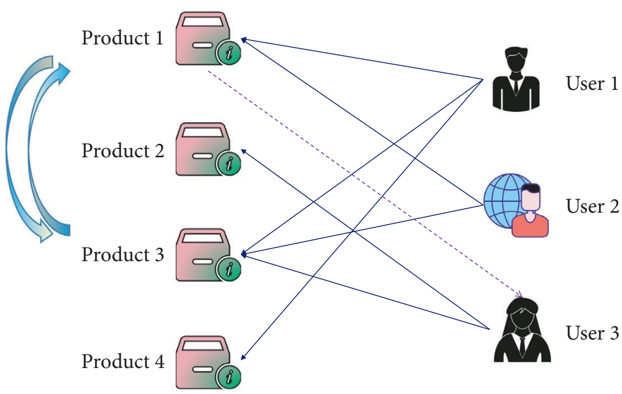


FIGURE 1: Schematic diagram of project-based CF algorithm.

classroom teaching and enhances the effectiveness of classroom teaching as well as it plays an extremely important role in the ideology course to implement the fundamental task of cultivating people by virtue.

The core idea of "precise teaching" in ideological and moral course is to meet students' individual differences, show flexibility in teaching, teach students in accordance with their aptitude, and give full play to the mainstay role of ideological and moral course in moral education. Teaching objectives play a very important role in the teaching process. Teaching activities are not only guided by teaching objectives, but also closely focused on achieving teaching objectives. Whether or not the teaching objectives are realized, and the degree of realization depends not only on the development of teaching activities, but also on the scientific and precise degree of setting the teaching objectives. Precision teaching is based on IT, and it is a typical "Internet plus education," which is helpful to accurately locate students. It truly reflects the actual situation of students in an objective and quantitative way, teaches students in accordance with their aptitude, respects individual differences, and conducts hierarchical teaching to meet the needs of students' individualized development, improve the atmosphere of traditional classroom teaching, and enhance their attractiveness.

Teaching resources are an important part of the implementation of precision teaching, play a vital role, and are an important support for the development of teaching process. Fragmented teaching resources have the characteristics of small capacity, wide content, and flexibility, which can meet the fragmented learning needs of current students and enhance the richness, vividness, and extensibility of teaching content. With the support of IT, ideology teachers can actively optimize the teaching process and effectively integrate the data recording and analysis of students' learning into the teaching process. In addition, ideology teachers optimize the introduction of digital teaching resources related to ideology courses, provide personalized teaching resources for students, and enhance the attraction of classroom teaching. According to learners' learning styles, learning needs, and learning abilities, this article adds personalized recommendation function to the teaching system, thereby avoiding the blindness of students' learning. The model architecture is shown in Figure 2.

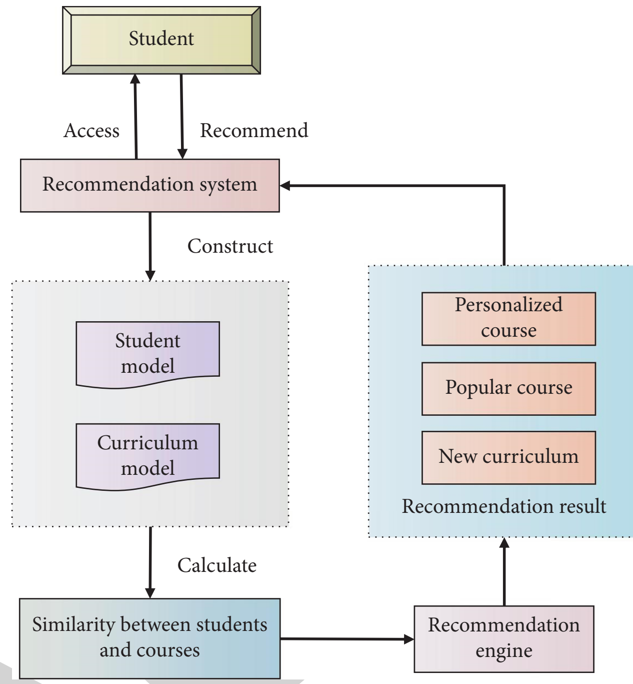


FIGURE 2: Student personalized recommendation model.

With the help of new IT, ideological precision teaching uses smart phones, tablets, and other intelligent terminal devices to make the classroom system transcend the limitations of time and space and realize more open classroom teaching activities, which is conducive to activating students' interest points and making students truly become the main body of learning. By studying the data analysis of students' learning behavior, teachers can predict students' future learning performance, adjust teaching strategies accordingly, and evaluate students comprehensively and objectively. Especially it makes the process teaching evaluation more accurate, so as to put forward feasible suggestions and countermeasures for students to achieve the expected teaching objectives and improve the effectiveness of ideology teaching. Educators should make accurate use of the existing teaching resources on the Internet. With the continuous development of network and IT, the available effective resources for ideology teaching are becoming more and more abundant. The construction of teaching resources for ideology courses should make targeted use of the existing network teaching resources and give full play to their advantages, so as to relieve the pressure of developing teaching resources independently.

Ideology courses in universities help students to build important courses of world outlook and outlook on life and values. Through theoretical study and practical experience, they can solve practical problems such as ideology, politics, morality, law, etc., that modern students encountered during his growing up. This requires the construction of ideology classroom to highlight the needs of students. Through various educational technology support, big data support, access to related platforms, etc., we can understand students' needs more accurately. Teachers can accurately supply teaching content by designing teaching topics, taking teaching materials as the basis, and taking common problems of higher vocational students as the guidance. Fragmented teaching resources are integrated into classroom teaching, and classroom teaching is carried out through the process of "looking at cases-discussing problems-talking about theories" and improve the informatization literacy of ideology teachers in universities. It is necessary to strengthen teachers' awareness of information-based teaching, enhance their attention and sensitivity to IT, and give full play to their subjective initiative. For the traditional ideology class, there are some limitations in the use of IT. Teachers of the two courses should actively learn the cutting-edge modern educational IT, effectively integrate IT with ideology class in universities, and reshape classroom teaching.

3.3. Precise Teaching Model of Ideology Courses Based on CF Algorithm. Through the analysis, it can be concluded that it is of research and practical significance to propose an accurate recommendation algorithm and build an accurate teaching model of accurate ideology course based on it. There are some problems in the cold start, sparseness, and expansibility of the traditional recommendation algorithm, which cannot be applied to the precise teaching of ideology courses. At the same time, some data can supplement the recommendation results, and these data are not fully utilized in traditional recommendation algorithms. Aiming at the data sparseness and cold start problems of CF algorithm, it can be solved by the attributes of items, namely CF recommendation based on item attributes and CF recommendation based on attribute value preference matrix. CF algorithm calculates the user's rating of the project as data, and then obtains the recommendation result. In the calculation process, the content of the project is not considered, which is very suitable for many educational data. It is difficult to determine the complete meaning of some educational data by keywords. If content-based algorithm is used for recommendation, it may lead to inaccurate recommendation results. The recommendation process based on this collaborative filtering algorithm is shown in Figure 3.

The data layer is mainly classified and stored according to the corresponding data constraint relationship based on the running requirements of personalized recommendation system. The data layer consists of personal information database, interest database, self-test question information database, and personalized information database. In education, there are huge educational resources. If we define the content of each educational resource, we need to spend a lot of time and energy. The recommendation result of CF algorithm to users is the k items that they are most interested in, and the quality of this recommendation method is higher. Moreover, the CF algorithm can also recommend potential items of interest to users, not just the items related to their rated items.

The similarity calculation in collaborative filtering algorithm is based on the data that each user scores at least two or more items together. Applying collaborative filtering algorithm to the precise teaching of ideological and political courses can accurately grasp the teaching objectives and contents and build a scientific teaching process. No matter what kind of recommendation is made, it is necessary to calculate the user's rating on a certain item, and the prediction result can be obtained by using the following formula:

$$r_{ui} = \frac{1}{k} \sum_{w \in N} r_{wi}, \tag{1}$$

where r_{ui} is the rating of item i by user u. N represents the calculated neighbor set. w denotes a neighbor of $u's$ n neighbors that has a rating for item i. k represents the number of all neighbors of u that have ratings for item i among the n neighbors of u. The final prediction result is obtained by unifying the criteria of all users and then weighting. It can be calculated by the following formula:

$$r_{ui} = \overline{r}_u + z \sum_{w \in N} sim(u, w)\left(r_{wi} - \overline{r}_w\right), \tag{2}$$

where z is the normalization factor, which is expressed as follows:

$$z = \frac{1}{\sum_{w \in N} sim(u, w)}, \tag{3}$$

where and is the average score of the product by the predicted user u and its neighbor user w. sim(u, w) represents the similarity between user u and user w.

The application layer mainly embodies the process of intelligent recommendation. The main function is to collect information about students' browsing, learning, and testing. According to the personalized recommendation algorithm, several knowledge points that are most similar to the target students are calculated, and the knowledge points are recommended to the target students, mainly including the student information collection module and personalized recommendation module. An evaluation matrix is established in the precise recommendation system to describe the main factors of students' course learning process, such as hobbies, majors, learning level, course selection records, and teacher evaluation. The algorithm analyzes students' behaviors according to the above lines information, establishes corresponding student items, finds out the course selection records with the highest similarity by comparing with the items in the evaluation matrix, and recommends courses to the students.

Statistical accuracy is generally calculated using mean absolute error (MAE) and mean square root error (RMSE). MAE calculates the deviation between the predicted score and the actual score by the absolute value. The smaller the deviation value, the more accurate the prediction. Assuming that the user's rating set for items is $\{P_1, P_2, \ldots, P_n\}$, and the user's predicted rating set for unrated items is $\{Q_1, Q_2, \ldots, Q_n\}$, then MAE is expressed as

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \left|Q_i - P_i\right|}{n}. \tag{4}$$

The smaller the MAE value, the higher the recommendation quality. Suppose $T(u)$ is a list of user actions on the test set. $R(u)$ is a list of recommendations made to users

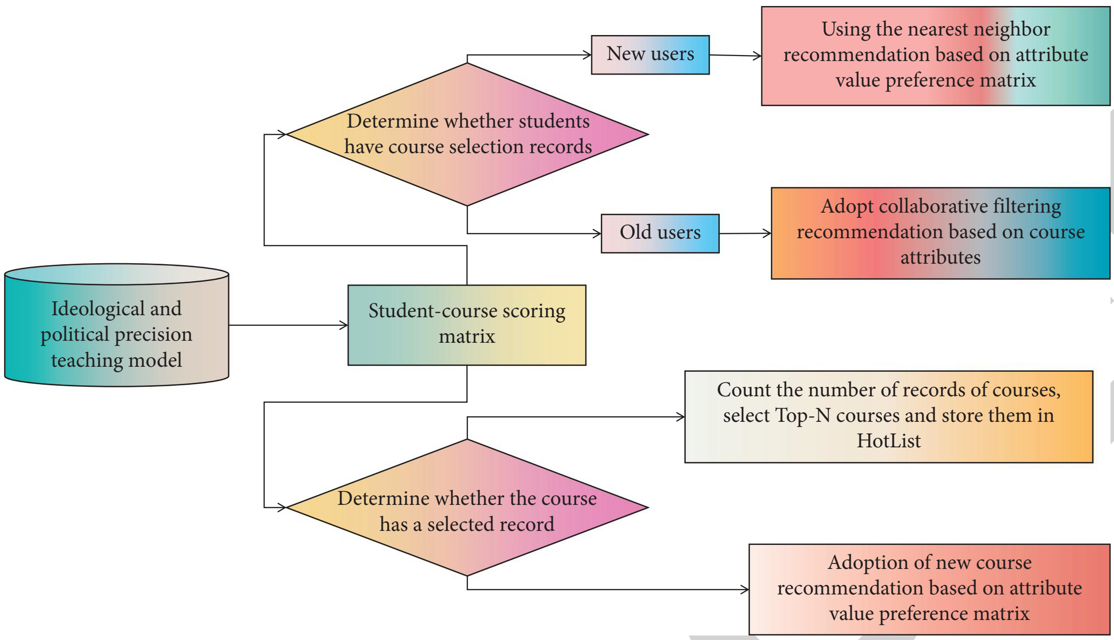


Figure 3: Recommendation process of collaborative filtering algorithm.

based on their behavior on the training set. The calculation of its precision rate and recall rate is shown as follows:

$$\text{Recall} = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |T(u)|}, \quad (5)$$

$$\text{Precision} = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |R(u)|}. \quad (6)$$

At present, the main evaluation indexes of recommendation algorithm are MAE, accuracy, and recall rate. According to the above three indicators, this article analyzes the recommended algorithm of ideological and political precision teaching.

The attributes of items or users are determined according to the relationship between users and items and themselves in the personalized recommendation system. Their attributes can reflect the uniqueness of items and users. In order to recommend any item in the recommendation system to the appropriate users in real time, we should first obtain the unique data of the item or user and construct the attribute vector space of the item or user, which will help to recommend the items that users need or like according to the CF algorithm. At the level of algorithm, because different tags of a project have different recognition degrees, it is necessary to measure the weight of different tags in actual calculation to calculate the correlation degree between goods and different tags. At the same time, due to the fuzziness and approximation of labels, some methods should be introduced to deal with labels to eliminate or reduce the influence caused by the fuzziness of labels. In the aspect of engineering development, we need to develop an easy-to-use evaluation and recommendation module. On the premise of obtaining data, these two modules must reduce the operation process of users as much as possible.

The vector-based method is used to calculate the similarity quantitatively. Based on the Euclidean distance calculation theory, it is assumed that x and y are any two points in the n-dimensional space, and the Euclidean distance between them is shown as follows:

$$d(x, y) = \sqrt{\left(\sum (x_i - y_i)^2\right)}. \quad (7)$$

According to the above formula, when $n = 2$, the Euclidean distance is the distance between two points on the plane. In order to use Euclidean distance for quantitative calculation of similarity, formula (7) is transformed as expressed below:

$$sim(x, y) = \frac{1}{1} + d(x, y). \quad (8)$$

In the above formula, the distance between x and y can represent the quantitative preference of students for a certain ideology course.

Based on the information of neighboring students and neighboring ideology courses, the calculation recommendation mechanism is formed, that is the preference of all target students for a certain ideology course is taken as a vector to calculate the similarity between students and ideology courses, and after the similar courses for a certain ideology course are obtained, according to the historical preference of target students, the current ideology courses that students have not expressed their preference are predicted, and a list of ideology courses arranged in a continuous and orderly manner is calculated as the recommendation and prediction list. The similarity calculation method of tags in the algorithm is based on the following theory: if two tags are marked on the same item, there is a certain relationship between the two tags, which is called “tag co-occurrence.” That is, the semantic relationship between tags can be obtained from tag co-occurrence. Tags clustered into tag clusters represent different topics, and each topic represents different semantics. In addition to the semantics of the tag itself, the tag of the item also has information such as tag times. For an item, if some of its tags come from a tag cluster, it is considered that the item is related to this tag cluster, and the proportion of tag tagging times to total tag times can quantify the correlation between the item and the tag cluster.

The algorithm can retrieve knowledge points that students may not have mastered and obtain the most "interesting" knowledge point set B according to the preset recommended number N of knowledge points or the similarity threshold. Finally, the union $A \cup B$ of set A and set B is recommended to the target students.

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{ic} - \overline{R}_c)(R_{jc} - \overline{R}_c)}{\sqrt{\sum_{c \in I_{ij}} (R_{ic} - R_c)^2 \sum_{c \in I_{ij}} (R_{jc} - R_c)^2}}. \quad (9)$$

In the precise teaching of ideology course, with the continuous learning of students, some relevant behavioral data will be generated. These data are the scientific basis for providing personalized learning and multi-evaluation learning services. Compared with traditional teaching evaluation, the most prominent feature of teaching evaluation in ideology classroom is its immediacy. With the help of the Internet and big data, it can timely analyze and evaluate the teaching process, especially the students' "learning" situation, so that teachers can constantly make adjustments in the subsequent teaching. Accurate learning evaluation of teaching makes up for the disadvantages of traditional evaluation, such as one-sidedness and lag, and realizes timely, effective, and accurate digital learning evaluation through the application of IT. The development of modern educational technology has the characteristics of the times. In order to closely follow the development trend of the times and change the traditional ideological and political classroom, education departments and schools should organize and carry out special training on information-based teaching, help school teachers master solid theoretical knowledge of modern educational technology, and strengthen the learning of information-based teaching technology.

4. Result Analysis and Discussion

The CF recommendation algorithm is used to retrieve knowledge points that the target students may not have mastered. The idea of the algorithm is to generate intelligent recommendation based on the interdependence between users: predict the "preference" of the target user according to the "preference" of neighboring users. The basic principle of limiting the maximum value of the students' neighbors of ideology courses is to ensure that all points in the area with the current point as the center and the distance K are used as neighbors of the current point. The algorithm can obtain an uncertain number of students' neighbors of ideology courses, but there will be no large fluctuations and deviations in the degree of quantitative preference, especially when dealing with isolated points. It improves the consistency of neighbor calculation in learning ideology courses. In the early stage of data mining, the original data are preprocessed to meet the experimental standards. The educational data used in this article come from the mobile independent school support platform. The data in the original database include all kinds of students' behavior data, which need to be filtered to screen out students' problem-solving records, and the final scoring matrix is obtained through implicit scoring. In the system data set, when the constraint index of tag similarity is 0.25, the adjustment coefficient in the similarity calculation formula is 0.4, and the number of nearest neighbors is 15, the system precision recommendation engine can achieve the best recommendation effect. In order to study the particularity of educational data, this article experimented on MovieLens data set and finally compared the experimental results of the two to analyze the particularity of educational data.

In order to verify the actual working effect of the precise teaching model of ideology courses established in this article, we conducted several experiments. The educational data are randomly divided into 10 parts according to a uniform distribution: 2 parts are selected as the test set and the remaining 8 parts are used as the training set. Then, the user's interest model is obtained through the training set, and the corresponding prediction is given on the test set according to the user's interest model, and an appropriate evaluation method is selected to calculate the recommendation effect. Figure 4 shows a comparison of the obtained accuracies for different kinds of recommendation algorithms with different values of k.

As can be seen from the figure, in the education data, the accuracy of CF algorithm based on users is low. This is because the number of students in the educational data in this article is much smaller than the number of exercises, and the sparsity of the data is also very large. When the user-based CF algorithm is used to find neighbors, there will be deviations, which will lead to the decrease of accuracy.

The functional module design of the system follows the principles of practicality, modularization, and expandability. The core modules of the system mainly include the submodule of: selecting courses for ideology courses in universities, evaluating students for ideology courses in universities, recommending courses for ideology courses in universities, and maintaining and updating the system. Each submodule works cooperatively under the control of the system workflow: constructing an efficient and practical closed-loop dynamic recommendation mechanism for ideology courses in universities to form a virtuous circle, providing basic guarantee for the development of ideology education in universities, and getting the basic information of students by registering information in the system. Then, through the learning style measurement scale and course selection records filled in by students, the learners' personality preferences and academic information can be obtained, so that the information of students can be mapped to the student model more accurately.

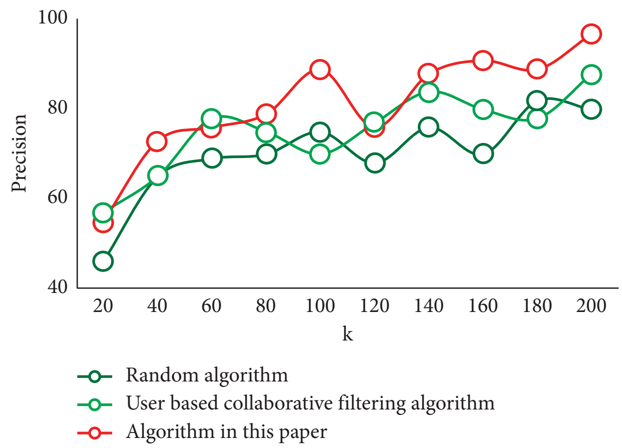


Figure 4: Accuracy of different algorithms under different *k* values.

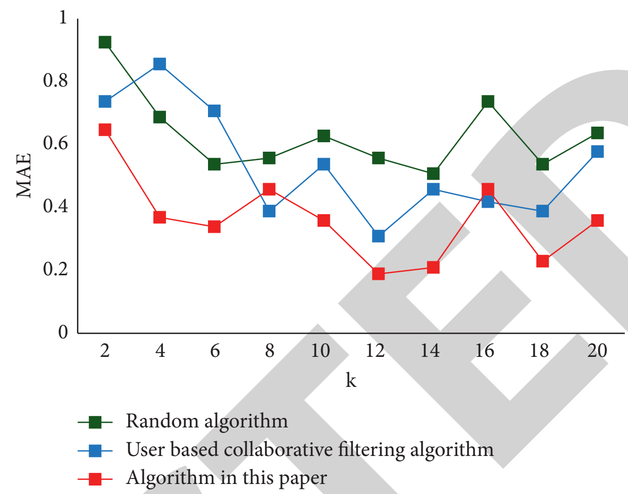


Figure 5: Line chart of MAE value under different algorithms.

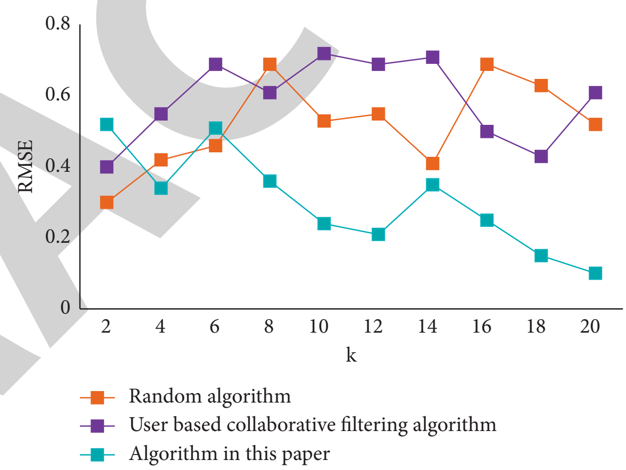


Figure 6: Line chart of RMSE value under different algorithms.

After determining the application student group, the system initialization operation is carried out. It mainly completes the input of the information of ideology courses currently offered by the school and students' preference information for each course in the past historical cycle. The initial value is written into the system data warehouse as the initial cold start data set of the improved CF algorithm. According to the student model and curriculum model in the system, the similarity among students, between courses, and between students and knowledge points is calculated, and suitable recommendation engines are used to recommend suitable content for students because the system includes students who log in for the first time and students who log in for more than two times and new courses. By comparing the project-based CF algorithm with the user-based CF algorithm, the recommendation effect of the precise recommendation engine of this system is verified, and the test results are obtained. MAE values of different algorithms are shown in Figure 5. RMSE values of different algorithms are shown in Figure 6.

The MAE and RMSE values of the three algorithms can be visually compared by the trend chart. It can be seen that with the increase of the number of nearest neighbors, the MAE value of each algorithm first decreases and then increases. Overall, the MAE and RMSE values of this method are the lowest, while those of the other two methods are slightly higher. This verifies the effectiveness of this method.

In order to comprehensively test the recommendation effect of the intelligent and precise teaching system of ideology courses, students are organized to try out the system. First of all, two sets of test questions are stored in the test library of the system, each set of test questions has 20 multiple-choice questions, and the difficulty coefficients of the two sets of test questions are equal and the test sites are the same. The users of the system interact with other layers through the user application layer and the extended service layer. The system determines the user type according to the information submitted by the users and decides whether to provide common user interface, teacher interface, or administrator interface for the users. Some operations related to the course are handled by the administrator in the background of the system, and the results of the course are entered by the teacher through the teacher interface. The optimized model is simulated and verified from the recall rate of precise teaching of ideology courses. The final simulation result with experimental data set is shown in Figure 7.

It can be seen that the optimized algorithm in this article has a higher recall rate, and its performance is improved compared with the traditional algorithm. In the experiment, the user interest data set and the user invisible interest data set are regressed and mapped.

It is necessary to start the submodule of personalized recommendation of ideology courses systematically and make personalized and accurate recommendation of ideology courses for different students so as to improve students' interest points and ensure that ideology in colleges and universities forms a three-dimensional education situation of "watering flowers and roots, teaching people and teaching their hearts". The learning forum provides a platform for communication and cooperation in the teaching mode of autonomous learning and cooperative learning, which can solve the problems that students

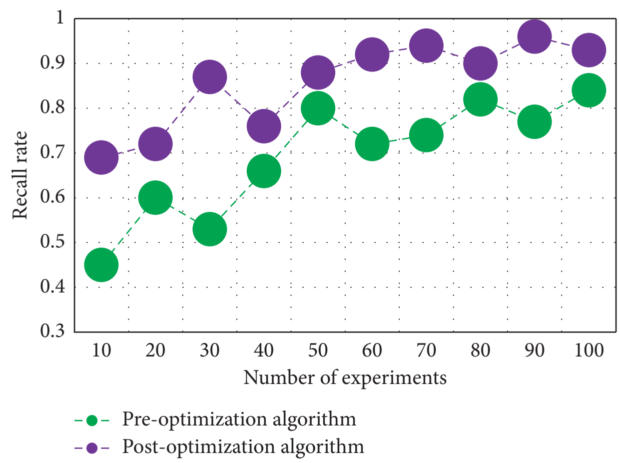


Figure 7: Comparison curve of recommended recall rate of the two models before and after optimization.

encounter in the course of learning. For the problems in study and life, we can also communicate with students and teachers through this platform, thus enhancing the friendship between students and teachers and students.

Hundred students are randomly selected to test the first set of questions, the system produces recommendations for different students according to the test results, and the students who organize the test identify the recommendation results produced by the system. The purpose is to confirm whether the knowledge points recommended by the system are not mastered and understood by the target students, that is the accuracy of the target students' identification of the recommended results. According to the recommended results of the system, the students comprehensively review the knowledge points, and after the review, the students are organized to test the second set of questions to check whether the students' scores have been greatly improved. This article uses the precision teaching system to experiment its practicability and effectiveness. The students' grades are checked for improvement and obtained the results as shown in Figure 8.

It can be seen that after using this method, students' scores have been significantly improved. The course evaluation module provides users with a platform to evaluate courses. Users can score the courses they have taken, such as evaluating a course from the teaching design and content of the course, the teaching methods and teaching effects of substitute teachers, teachers' ethics and academic characteristics, etc. In order to adapt to the variability of students' scoring information in the time dimension, the historical preference fusion similarity set is extended to a dynamic data set. Gradually, the actual needs of dynamic changes of interest points of college students in the new period are adapted, and the personalization, pertinence, and accuracy of recommendations are improved.

This article introduces the framework of the precise teaching system of ideology courses. The data generated by the teaching system is used for experimental test, and the related parameters are adjusted according to the experimental results and compared with other algorithms. The

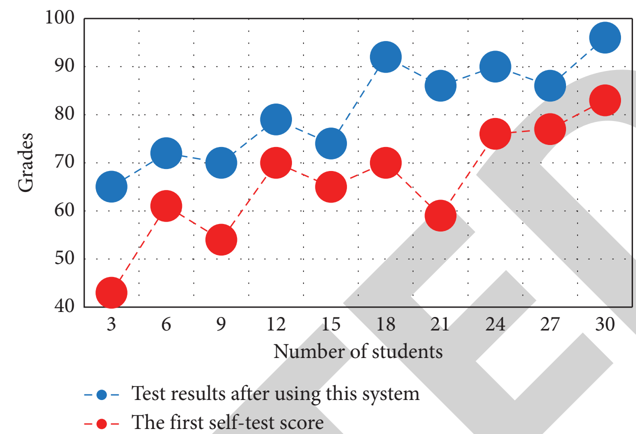


Figure 8: Comparison line chart of students' two test scores.

analysis of the experimental results shows that the precision recommendation engine constructed by the algorithm proposed in this article has improved the precision teaching effect to a certain extent, and the performance of the improved algorithm is also improved. Therefore, if the CF algorithm can be applied to the current teaching system, it can not only provide convenience for students to check for missing items, but also let teachers know how well each student has mastered the course.

5. Conclusions

As a public compulsory course, most schools teach in large classes, which easily lead to teachers' low attention to students in class and cannot meet the individualized development needs of students. Classroom teaching ideas are old-fashioned, students' acceptance is not high, and it is difficult to achieve accurate teaching, so it is difficult to achieve ideal teaching effect. In order to better meet the three-dimensional educational development trend in ideology education in universities under the new situation of "watering flowers and roots, teaching people, and teaching heart," an atmosphere of ideology education in universities with win-win cooperation between teachers and students, diverse forms, and innovative personality is created actively. In this article, aiming at the problems existing in the curriculum of ideology education in universities, a recommendation system of ideology education in universities based on improved CF algorithm is constructed and the goal of precise teaching of ideology courses is achieved.

Precise teaching organically integrates the requirements of ideology course, the characteristics of higher vocational education, and the needs of students, taking into account the three needs in the setting of teaching objectives, the construction of teaching resources, the design of teaching process and the reconstruction of evaluation system, and accurately docking, effectively solving the problem that the pertinence of ideology course teaching in schools is not strong. The teaching system of this article is not only an important tool to realize the efficient and accurate teaching management of ideological and political courses, but also very beneficial to strengthen students' quality education and

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9807328, 1 page
https://doi.org/10.1155/2023/9807328

Retraction

Retracted: Emotion Analysis for Foreign Language Learning under Scenarios of Network

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert reader that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zheng and F. Xu, "Emotion Analysis for Foreign Language Learning under Scenarios of Network," *Security and Communication Networks*, vol. 2022, Article ID 2789291, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2789291, 9 pages
https://doi.org/10.1155/2022/2789291

Research Article

Emotion Analysis for Foreign Language Learning under Scenarios of Network

Jingzhi Zheng[1] and Fang Xu[2]

[1]*School of Foreign Languages, Hubei University of Science and Technology, Hubei 437000, Xianning, China*
[2]*School of Foreign Languages and International Education, Wenzhou University of Technology, Wenzhou 325000, ZheJiang, China*

Correspondence should be addressed to Fang Xu; xuf@wzu.edu.cn

Received 28 March 2022; Revised 23 April 2022; Accepted 27 April 2022; Published 11 May 2022

Academic Editor: Chin-Ling Chen

With the development of the Internet, modern educational technology is also making continuous progress based on the Internet. Internet technology provides us with abundant online learning resources and creates a brand new teaching mode. Changes in the field of education bring new opportunities and challenges, which affect students' mental and physical health. This study focuses on the psychological problems of students under the background of network teaching. Through data analysis, it can be concluded that emotions not only affect students' interest in learning but also affect their learning process. The results show that emotion plays an important role in students' learning, and different sources of learning emotion and emotional experience have different effects on students' academic performance. Teachers must scientifically use modern educational technology to establish a good learning environment and, at the same time, help students to have a healthy psychology. Only having a healthy body and mind can improve one's comprehensive ability.

1. Introduction

With the rapid development of information technology, great changes have taken place in all aspects of society, and the world has entered the Internet era. Under the influence of the Internet, modern educational technology has changed greatly, and online learning, as a new educational model, has become a popular way of learning [1]. With the support of information technology, online learning breaks the traditional teaching mode and makes the multi-interaction between students, teachers, and network learning resources across time and space become a reality. In the context of modern educational technology, network education platform enables intelligent and in-depth communication and interaction between teachers and students and computer learning system [2]. In this era, the exchanges between different nationalities in the world are also becoming increasingly frequent. Language, as the basic medium of communication and the carrier of cultures, is particularly important. Only by understanding the language can we understand the culture of a country. In this context, the importance of foreign languages is very prominent. Modern educational technology continues to reform and develop, among which the biggest impact is foreign language learning. Foreign language learning not only needs a lot of materials but also needs to understand the cultural knowledge of the foreign language so that learners can easily find their own learning methods. The online platform provides a large number of learning materials for foreign language learning, and students can study online freely and independently [3]. At the same time, with the expansion of network influence, students' learning psychology will also have a certain change. Therefore, educators must pay attention to students' mental health.

Studies on college students' learning focus on the field of mental health, and studies on individual learner differences mainly focus on intelligence and linguistic ability, while studies on emotional factors are ignored. However, it is far from enough to explain the particularity and complexity of foreign language learning only from a cognitive perspective [4]. With the development of humanistic psychology, the focus of language teaching has gradually shifted from how teachers teach to how students learn, resulting in more and more researchers' attention to emotional problems in foreign language teaching. The affective filtering hypothesis of Krashen [5], an applied linguist, holds that emotional barriers prevent language learners from fully obtaining comprehensible input. When the emotional barrier is strong, language learners may be able to understand the language input, but there will be a screening of language materials, and all materials cannot be smoothly delivered to the language acquisition device for internalization. If the barriers of emotional factors are smaller, the language input will proceed more smoothly and the foreign language learning effect will be better [6]. In this context, researchers are paying more and more attention to the influence of emotional factors on foreign language learning. Due to the examination pressure and learning burden caused by the exam-oriented education system, learners are becoming more and more anxious. Foreign language (English is the most popular foreign language in many countries) has become a subject with a higher level of anxiety among many subjects. The cultivation of emotion in foreign language has become one of the hottest topics for scholars at home and abroad because it is not only an emotion factor but also an important variable in the study of "individual differences." Therefore, it is necessary to study the emotional factor of foreign language anxiety. In order to arouse students' enthusiasm for learning a foreign language and enhance their ability of foreign language learning, foreign language teachers should pay more attention to the complexity and particularity of foreign language learning and attach importance to emotional factors in foreign language learning [7]. On this basis, they should flexibly use teaching methods to effectively organize foreign language teaching activities and establish a harmonious teacher-student relationship, which can effectively relieve students' foreign language anxiety to a certain extent and encourage students to master the method of regulating emotions through Internet technology. Therefore, this study takes college students' mental health as the research object to study the status quo and educational countermeasures of their foreign language learning in online environment. Through specific scientific data analysis to prove the correctness of the research, find the specific emotional state of students.

2. Related Work

2.1. Research on Emotional State of Student Learning. The most important function of human language is the communication between people, which will involve emotions definitely. People need to express their feelings, communicate with others, and solve problems with languages [8]. Foreign language, as a nonnative language to learn, is bound to cause problems and setbacks to students in the process of learning. Therefore, students with a positive emotional attitude will be of great help to learn English well. Emotional factors are also a part of psychological quality, which is an important ability of contemporary college students to adapt to the fierce competition of modern life. Therefore, to pay attention to the emotional factors of college students, which reflects the educational thought of taking students as the main body, emphasizing, understanding, and respecting the individual needs of students and paying attention to the cultivation of students' quality are of great significance to cultivate students' ability of independent learning [9].

The development of information technology began to enter the Internet era. A series of information technology synergistic effects have triggered a profound social impact, resulting in many fields including philosophy, psychology, and education, and began to use this new technology. The application of network technology in these fields has long attracted the attention of scholars from all walks of life abroad. Dulay et al. proposed the "affective filter hypothesis," which made a preliminary study on the affective factors related to the learning process, in order to explain how affective factors affect foreign language learning process. Later, Krashen developed the theory on this basis, and he believed that affective filtering is a psychological disorder that prevents learners from completely digesting the comprehensive input obtained in their learning [5]. According to the affective filter hypothesis, language learning varies from person to person, and the differences are mainly due to psychological reasons because each learner has a different intensity in learning motivation, personality, and attitude, thus forming corresponding strong or weak psychological barriers. Sha believes that the more motivated learners are, the more confident they are, the less anxious they are, and the less they filter language input, and the more input they get, the better their learning results will be. On the contrary, learners with low motivation, low self-confidence, and anxiety absorb less comprehensible input and thus have a poorer effect in foreign language learning. Morris distinguishes between positive and negative emotions [10]. He believes that when we study the emotional factors of language learners, we should focus on how to overcome the problems caused by negative emotions and how to create positive and beneficial emotions. He believes that learners' emotional state directly affects their learning behavior and results. Ren concluded from his own and others' research that emotions play a decisive role in foreign language learning and that any failure to learn a foreign language can be attributed to a variety of emotional disorders [11].

2.2. Emotion of Foreign Language Learning under Network Technology. Online learning environment is a necessary condition for the introduction of network factors into the learning situation and the development of learning activities [12]. The learning psychology of learners under the influence of modern educational technology is obviously different from that in the traditional learning environment [13]. The uniqueness of learning mode in the network environment will lead to the uniqueness of learning psychology; especially, learners will have obvious feelings of loneliness and anxiety. Supporting asynchronous learning is one of the most celebrated advantages of online learning, which breaks the limit of time and space for learning activities. However, from another perspective, the separation of time and space between teachers and students will cause a series of problems. For example, the uncertainty of learning situation and the lack of "real contact" will cause anxiety among learners. Educational research under the network environment needs to combine technology, psychology, and teaching [14]. NaoKo, a Japanese scholar, believes that the psychological basis, namely, emotion, is the core of autonomous learning. Many researchers have realized that emotional factors can affect the effect of learning, but there are few studies that combine emotional factors with autonomous learning. For example, English, as a foreign language, should be learned with a positive emotional attitude by learners [15], so as to mobilize their enthusiasm and initiative of learning, thus cultivating the ability of independent learning. Therefore, emotional factors are the key to cultivate the independent learning ability [16].

With the further study, researchers gradually found that emotions play a very important role in online foreign language learning [17], among which network technology factors, teaching process, and online course guidance all affect students' emotions. For example, the delay of online system will make students feel nervous. Researchers have found that, in online foreign language learning, when learners' goals are very clear, they will establish a learning direction, manage learning goals, and improve learning motivation. Meanwhile, learners' time management ability, emotion management ability, their own learning foundation, and learning support are all factors that affect students' online learning ability [18]. To some extent, the emotions generated by teachers in the process of online teaching as well as how to choose teaching content, design corresponding online teaching activities, and supervise and guide students' learning process are related to the determination of teaching objectives. All these aspects will affect teachers' emotions [19]. In online foreign language teaching, teachers can skillfully drive students' learning atmosphere to feel relaxed and promote students to develop healthy learning emotions. Therefore, whether teachers can create a healthy learning emotional atmosphere plays a key role in students' learning.

3. Research Design

3.1. Emotion of Foreign Language Learning Based on Network. With the progress of society, we should not only care about students' grades but also know their mental health. Most of the research studies on foreign language learning under modern educational technology are concentrated, including teaching methods, teacher feedback, and technology-related aspects. Researchers in the field of computer mostly conduct research studies on the influencing factors of online foreign language learning from the perspective of technology. Network learning is a new form of online information transmission with the rapid development of Internet technology [20]. Because of the development of online learning technology, students' learning environment has changed. Online foreign language learning makes full use of the characteristics of Internet technology, such as powerful online interaction, no geographical restrictions, and simplicity, promoting the popularity of foreign language learning. More and more students are carrying out online live streaming of education activities, and the research on all aspects of online learning environment is gradually carried out [21].

Our method is based on the process, causes of students' emotions, and considers the main environment in which emotions function. Firstly, this paper studies what academic emotions college students will have in the process of online foreign language and what aspects these academic emotions come from. Three hundred and fifty college students were randomly selected as interviewees, including 150 boys and 200 girls. The basic information of interviewees is shown in Table 1.

This research will use qualitative method based on grounded theory (grounded theory is the process of collecting and analyzing relevant data for a phenomenon system and then discovering, developing, and testing theories from it [22]) to verify the academic emotions of college students in online foreign language classes and explore their sources. Text data are obtained through interviews with college students, and the interview content is analyzed by using the coding method of grounded theory. Qualitative research is an integrated exploration and study of certain social phenomena in a natural context with the researcher himself as the research tool and various data collection methods, analyzing the collected data by induction. The qualitative research method does not set a frame in advance, but carries out in-depth and meticulous interaction with the researcher in the research, so as to obtain the overall explanation of this phenomenon.

The interview method is a method to obtain original data through oral communication with the research object. As the most commonly used specific research means to collect data, the interview method uses research tools such as recording pen, interviewee's consent form, and notebook to ensure the complete use of information in interview. Considering the purpose of the study is to investigate the academic emotions of college students in the online learning environment and the sources of academic emotions so as to prepare interview questions, college students who have attended network classroom were interviewed and the interview content was recorded. Afterwards, Word and qualitative analysis software Nvivo12 were used to transcribe and encode the interview materials. The interview questions are structured so that students can freely express their real thoughts. A small number of students who could not be interviewed face to face in formal interviews were interviewed online through Wechat, Dingding, and other communication media with their consent. Most students mainly conduct face-to-face interviews, and in the process of communication, they can also pay attention to the nonverbal responses of interviewees. The researchers recorded the content manually during the interview or recorded the interview with the consent of interviewees. During the interview, the researcher maintained a respectful attitude and listened carefully to the

Table 1: Basic information of interviewees.

Major	Man	Woman	Total
Math	20	40	60
Computer	20	30	50
Electronics and information engineering	30	10	40
Language and literature	10	20	30
Psychology	30	30	60
Ethnonymics	10	30	40
Biology	10	30	40
Art and design	20	10	30
Total	150	200	350

interviewees' answers to the questions. They used Word to transcribe and saved the recordings after the interview. In this study, the minimum interview time of a single interviewee was 10 minutes, the longest was 1 hour, and the average interview time of each interviewee was 30 minutes, with a total of 100,000 words of transcript. In order to ensure the personal information security of interviewees, their information is blurred. It is easier to find students' real emotions in face-to-face communication and survey.

3.2. The Influence of Web-Based Learning Emotion on Learning Effect. Based on the research of the first stage, the main academic emotions experienced by college students in online live teaching are happiness, relaxation, boredom, and anxiety. Considering that academic emotions mainly come from two aspects, teaching method and teaching environment, we further investigate the impact of emotional sources and emotional experience based on online foreign language learning on the learning effect, as shown in Figure 1. More than 1,800 college students from different majors were randomly selected, including 780 male students and 1050 female students. According to the emotional assessment questionnaire used by Gross and Levenson (1995), in the study, the emotional assessment of college students adopts 9 points, with 1 indicating that they do not feel such emotions at all and 9 indicating that they feel very strongly about emotions. Four emotional adjectives (happy, relaxed, anxious, and bored) were selected as the main content of the study according to the emotional experience induced by the experiment. We selected different types of online classroom contexts to investigate the emotional state of students in each situation.

3.2.1. Relax

(1) Teaching Environment. In the online foreign language course, the form of class is very flexible. The teacher does not require every student to turn on the microphone and the camera. You can choose the environment you feel suitable or accustomed to. You can lie on the sofa in your pajamas at home, with your favorite snacks in front of the computer. You are familiar with the use of network class. When students actively answer the questions raised by the teacher in the form of bullet screen, the teacher presents the feedback in the form of emoticons. For good speeches, the teacher gives thumbs up on the interactive panel. The pattern of teaching is fixed, and there will be no sudden mic or roll call. During the break, the teacher will play some music and organize students to do some interesting activities remotely.

(2) Teaching Method. In the online foreign language course, the teacher's teaching method is very flexible, creating a learning atmosphere. The teacher explains the knowledge point very thoroughly and naturally interacts with the students. There is no sense of distance between the teacher and the students when discussing problems together. The teacher has a comprehensive control of the knowledge content, and the theory learning will deepen the understanding of students with examples. After class, the assignment is moderate, and the teacher will give positive feedback to each student after completing the task. Teachers will extend a lot of knowledge points in the online class, and the feedback from teachers through the interactive panel is also very timely. You preview the knowledge to be learned today with the guide before class. You are very familiar with the knowledge points and can follow the progress of the teacher to study in an orderly way.

3.2.2. Happiness

(1) Teaching Environment. You can choose the way you like in online foreign language classes. There will be a break in the middle of the class, and the class schedule is compact and orderly. The teacher will ask the students to open the camera so that each student can see their own images, and their changes of expression and emotional will be more intuitively presented. Teachers are very familiar with the use of live broadcasting platforms and electronic devices. The network is smooth and the pictures of online classes are bright. The course mode is fixed, and the teacher will not suddenly check the mic or call the roll, but will communicate with the students on the interactive panel. The teacher will organize a group discussion and let students discuss freely on the Internet. Everyone can speak freely and the teacher will give feedback to each student in time on the interactive panel. In the online class, the course content is very clearly presented. At the same time, you can look up information anytime and anywhere. After the lecture, the teacher will use videos related to the course to impress the students. After class, you can also use the replay function to study repeatedly, check the omissions and fill in the gaps, and have a deep understanding of the knowledge points.

(2) Teaching Method. In the process of online foreign language courses, teachers will not follow the script, but explain the content based on their own experience or vivid and interesting examples. They will attract students' attention by talking about some topics they are interested in. The teacher prepares the lesson well before class and arranges the rhythm of the lesson. The teacher will not talk about too much digression, and the examples cited are very appropriate to help you understand the course knowledge more fully. Teachers communicate more with students and ask them to answer questions voluntarily. If the students give correct answers, teachers will praise them, fully recognize them, and accept

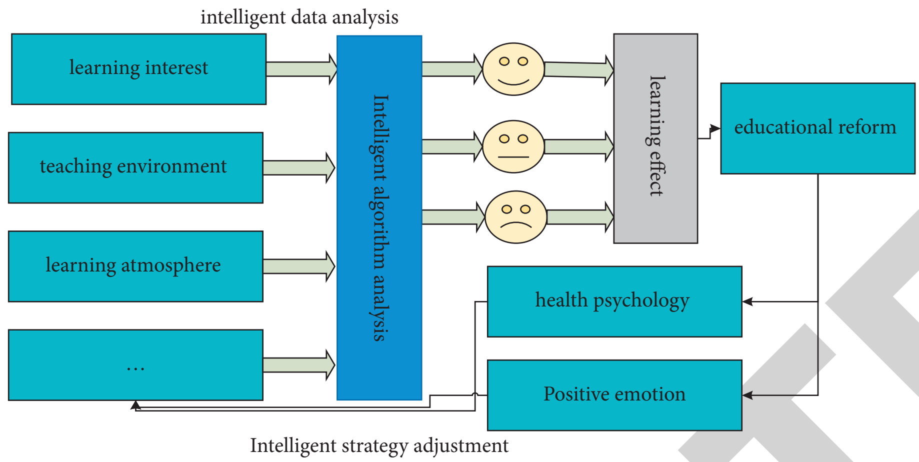


Figure 1: The relationship between students' learning emotion and learning effect.

diversified answers. If they give wrong answers, teachers will not criticize students, but give positive feedback and guide students to think in an appropriate way. Rude criticism will bring negative emotions and even psychological barriers to students and skillfully guide students to learn how to solve the problem is more important than the answer.

3.2.3. Boredom

(1) Teaching Environment. In online foreign language courses, teachers and students do not have cameras on so that they cannot see teachers and classmates. It is easy to be distracted all the time. In class, the teacher cannot see it when students apply for mic to ask questions, and no one answers questions on the interactive panel. If there are any questions or new ideas, the teacher cannot share them with students in time, and discussion and questioning will be very late. The teacher has explained a certain knowledge point in detail, but there are still students who apply for continuous mic and keep asking the teacher. When the teacher asks questions in class, no students apply for the mic, and no one answers the interactive panel. When the teacher is on the air, the picture is not smooth, and the electronic equipment always needs to be adjusted. Teachers have not mastered the functions of some platforms, and there are some technical problems; especially when playing video and audio, they need to constantly switch, which greatly delays the class time. Live broadcast time is long, and there is no rest in the middle, which will make the eyes particularly uncomfortable in class. Teachers, in order to meet class requirements, sometimes make the time of online classes too long. The environment of the teacher's network classroom is very noisy, and the volume of electronic devices is relatively low, which affects the class. Your home environment is not quiet enough, and there is no learning atmosphere. This objective learning environment is often the source of negative emotions.

(2) Teaching Method. In online foreign language courses, the teacher speaks too fast and the course schedule is too tight. You cannot follow and accept too many knowledge points since it is not easy to master. The teacher will talk about a lot of irrelevant topics and give inappropriate examples, which make it difficult to grasp the main points of the course. The content of the course is boring, and most of the time is spent on reading courseware. There are no videos related to the course during the whole class, and no activities are organized to adjust when students are distracted. In online class, due to the lack of interaction, some problems may not unfold. Some of the questions are already well understood, and they go back and forth on that question. The teaching mode is fixed and single, without innovation, causing students distracting and missing knowledge points. The homework assigned by the teacher after class is not given positive feedback and only used as the grading standard.

3.2.4. Anxiety

(1) Teaching Environment. In online foreign language courses, teachers are not familiar with the operation of computers and the use of software, so the video and audio conversion always wastes a lot of time. In an online class with no playback, the Internet is always not smooth, and you cannot concentrate because of the noisy sound of other students opening mics together. In online class, the surrounding environment is very noisy, which always disturbs you and makes you miss a lot of explanation of knowledge points. Because the network jam, the signal is not good, which makes the screen not clear and the courseware cannot be opened. There is no interactive intimacy because of the screen. In a closed space, there is no one to interact with you and give you feedback when you express your opinion. The class time is very long and there is no rest in the middle. You cannot focus on the class and find your harvest is very little. You forget to turn off the microphone, and your personal voice affects other classmates and teachers.

(2) Teaching Method. In online foreign language courses, the teacher's teaching progress is very fast and the explanation of knowledge points is not very thorough. You cannot keep up with the pace of the teacher, cannot accurately understand

the teacher's meaning, and miss the focus of the course. Your teacher will suddenly ask you a question online but you do not hear the question clearly and have no classmate to turn to for help. You do not know how to answer it without preparation. The knowledge points in the course are not consistent, and the key points are not very clear. It is easy to be distracted in class, and you would have missed a lot of content when you return to class. You cannot keep pace with teachers and classmates in learning and cannot timely consult teachers and classmates for the progress of the course and the knowledge points you cannot understand. In a word, negative emotions lead to a significant decrease in learning efficiency, which proves the importance of emotion in teaching.

4. Research Results and Analysis

4.1. Learning Emotion Based on Network

4.1.1. Analysis of Experimental Results. From the interview results, it can be found that the academic emotions experienced by college students in online foreign language learning are boredom, relaxation, happiness, anxiety, helplessness, frustration, tension, and anger. Through the analysis of the frequency and proportion of students experiencing academic emotions in online learning, it can be found that, about 94% of college students in the online foreign language learning experience boredom, 88% of college students in the network classroom experience to relax mood, 77% of college students would experience the happy mood, and 68% of college students experience anxiety. It can be seen that college students mainly experience four kinds of academic emotions in online classes, which are boredom, relaxation, happiness, and anxiety in sequence, as shown in Figure 2.

The emotional sources of college students in online foreign language learning are shown in Figure 3. The academic emotions of college students in online learning mainly come from teaching methods and teaching environment. The academic emotions of college students mainly come from the teaching environment, such as network environment, network broadcast platform, learning atmosphere, interaction between teachers and students, course participation, interaction between classmates, teacher feedback, and online resources. The academic emotions generated by college students also come from teaching methods: course organization, course time arrangement, knowledge framework presentation, and example analysis. In online foreign language learning, teachers' teaching style, difficulty of teaching content, and students' main factors are not the main sources of students' academic emotions.

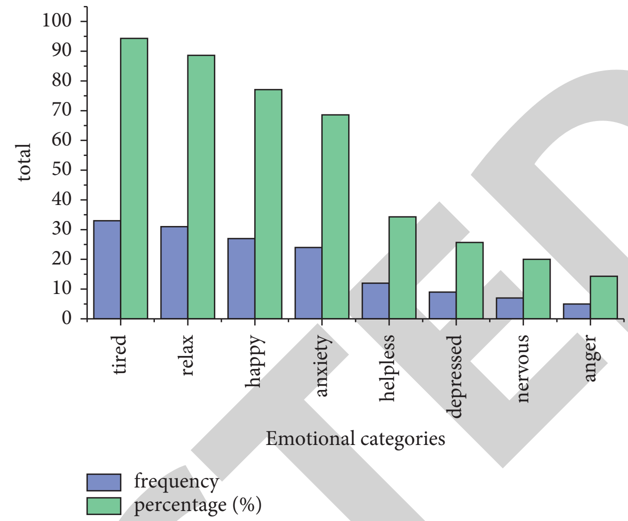


Figure 2: College students' emotions in online foreign language learning.

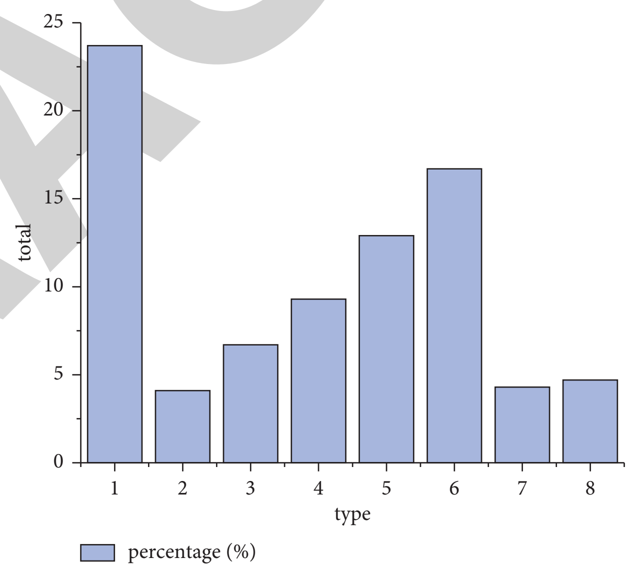


Figure 3: Emotional sources of foreign language learning under the context of modern educational technology (Notes:1, way; 2, time; 3, framework; 4, network environment; 5, live webcast; 6, learning atmosphere; 7, teacher's attitude; 8, students' attitude).

4.1.2. Discussion of Research Results. It can be found from the above research that the main academic emotions experienced by college students in online foreign language courses are boredom, relaxation, happiness, and anxiety. Emotional experience in online foreign language learning mainly comes from teaching environment and teaching method. Online foreign language learning provides a brand new learning environment. Students will have different emotional experiences when facing different network environments, live broadcast platforms, and online interaction modes between teachers and students. At the same time, teachers will also show different teaching methods in the process of live teaching. The organization of online class schedule, the arrangement of course time, and the presentation of relevant knowledge will give students different emotional experience. When college students are able to communicate with teachers at any time, carry out online group discussions with their classmates, and have high participation in the course, they will feel happy.

Teaching environment can make students produce positive academic emotions, such as good teaching atmosphere and harmonious teacher-student relationship, which can strengthen students' positive emotions. This is confirmed by the research results. In online foreign language learning, college students can interact with teachers and classmates in real time, participate in class discussions, share their own opinions, and answer questions directly with teachers online, experiencing happy emotions in a harmonious learning atmosphere. In the process of distance learning, teachers can accurately identify the emotional state of distance learners and timely give positive online feedback to students in the learning process, which can promote students to actively participate in the classroom environment, improve students' satisfaction with online learning, and effectively improve the online completion rate.

The phenomenon of network lag and delay will occur in online courses, and the unstable network environment will make college students feel bored. When teachers suddenly ask students to turn on their mics or cameras during online classes, students will feel anxious. Due to the diversity of teachers' teaching situations and students' learning situations, different situations and events in the learning process will affect the occurrence and change of students' academic emotional experience. Online academic emotional experience has a strong immediacy. The network environment has an important impact on the emotional experience of college students. Although the Internet develops rapidly in today's era, there are still some students in the learning environment without wireless network. In the network classroom, the stability of the network environment is guaranteed, the courses can be smoothly carried out, and teachers and students can interact online in real time, which will effectively reduce the boredom of students. In the process of teaching activities, students can get teachers' timely response and positive feedback interaction, increasing the understanding of teachers and students to each other. Teachers can really help students adjust their learning attitude by listening to students' feelings and giving feedback information seriously. The environment for students to take online courses is generally closed and lacks learning atmosphere. The learning environment of network classroom is virtual. When teachers and students have real-time dialogue in class, students' sense of participation becomes stronger. Besides, online discussion allows students to adjust their state and gradually focus on the learning atmosphere.

4.2. The Relationship between Modern Educational Technology and Emotion

4.2.1. Analysis of Experimental Results. Table 2 shows the mean and standard deviation of the influence of emotional sources and emotional experience on learning effect of college students in online foreign language teaching.

Xuexixiaoguo also calculated the mean and standard deviation of emotional sources and emotional experiences of students of different genders in online foreign language teaching. The statistical result of male students' learning accreditation was 3.561.58 and that of female students was 3.31.59. SPSS22.0 was used for 2×4 two-factor analysis of variance. The independent variable is the source of academic emotion (teaching method and teaching environment), emotional experience (happiness, relaxation, boredom, and anxiety), and the dependent variable is learning effect.

Table 2: Mean and standard deviation of the influence of learning emotion on learning effect.

Classroom mood	Teaching environment (M ± SD)	Teaching approach (M ± SD)
Happy	4.33 ± 1.9	4.38 ± 1.53
relax	2.96 ± 1.48	4.01 ± 1.35
boredom	3.13 ± 1.18	2.44 ± 1.08
anxiety	3.3 ± 1.26	2.86 ± 1.71

The results show that there is no significant difference between different genders and different grades of college students' academic emotions on learning results in online foreign language teaching ($p > 0.05$). In the online learning effect of college students, the main effect of academic emotion is not significant, $F(1, 175) = 0.003$, $p > 0.05$. The main effect of emotional experience was significant, $F(3, 175) = 9.95$, $p < 0.05$, and the interaction between the source of academic emotion and emotional experience was significant, $F(3, 175) = 3.28$, $p < 0.05$. A simple effect test was conducted on the interaction between emotional experience and academic emotion source, and it was found that there was significant difference between relaxation emotion and academic emotion source level, ($F(1, 175) = 6.22$, $p < 0.05$). When academic emotion comes from the teaching environment, the difference between the four emotions is significant, $F(1, 175) = 4.30$, $p < 0.01$. When academic emotion comes from the teaching method, there is significant difference between the four emotions, $F(1, 175) = 9.08$, $p < 0.001$. In other words, the source of academic emotion and emotional experience will have an impact on the learning effect of college students, as shown in Figure 4.

According to the experimental results in Figure 4, when the academic mood comes from the teaching environment, the learning effect under happy mood is better than that under bored mood, and the learning effect under happy mood is better than that under relaxed mood. When the learning emotion comes from the teaching method, the recognition score under happy mood is significantly higher than that under bored mood, and the recognition score under happy mood is significantly higher than that under anxious mood. When the academic emotion is derived from the teaching method, the recognition performance under relaxed mood is significantly higher than that under bored mood. Learning emotion under relaxed mood is more conducive to learning when it comes from teaching method, while learning effect is not good when it comes from teaching environment.

4.2.2. Discussion of Research Results. By studying the influence of different sources of academic emotion and emotional experience on the learning effect in online foreign language teaching, it is found that when the academic emotion comes from the teaching environment, the learning

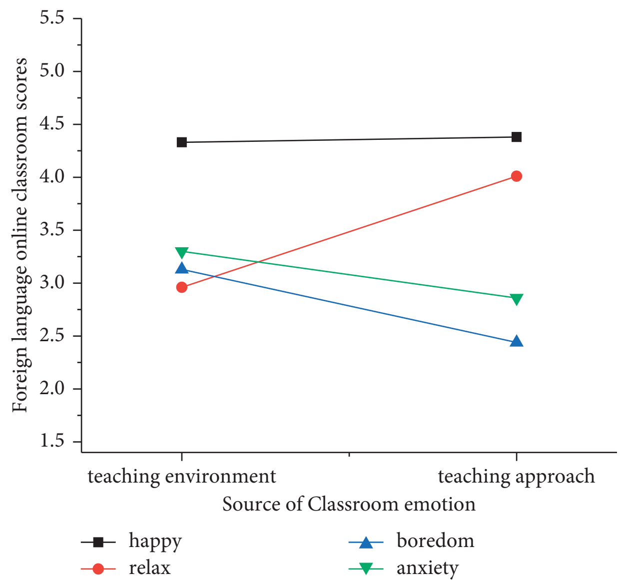


Figure 4: The relationship between classroom scores and emotional sources.

effect under happy mood is better than that under bored mood. Studies have shown that most students will be more motivated to learn when they are affirmed by teachers in their communication with teachers. It can be shown that teachers' affirmation is conducive to improving students' positive academic mood and promoting students' learning. Students can focus their attention in online classes, keep up with the pace of the teacher's course, and have a relaxed mood, so as to stimulate their interest in online learning, and the learning effect will be better. When the academic emotion under relaxed mood comes from teaching method, the academic performance under relaxed mood is significantly higher than that under bored mood. Through the network broadcast platform, the online platform is convenient and fast to operate, and the class atmosphere is strong when students communicate with teachers directly, and students have a stronger sense of participation. Students often use discussion boards and assessment tools to exchange ideas and receive quick feedback from their teachers. Students prefer to focus their learning on technology-supported collaborative learning, where interaction between students and faculty is enhanced by sharing knowledge.

When the happy emotion comes from the teaching method, the students' classroom performance is significantly higher than that under the bored mood, and the recognition performance under the happy mood is significantly higher than that under the anxious mood. In other words, the positive emotional experience generated by students in the process of learning is conducive to promoting the maintenance of students' interest in learning and can improve their interest in learning, thus improving their cognitive participation and learning effect. Online foreign language teachers need to give students pre-class guidance, mainly for course preview. In the process of online teaching, teachers set up corresponding discussion parts so that students can ask questions. When the number of students asking questions in class is small, the teacher can solve students' doubts in class. When a large number of students raise questions, teachers answer them one by one, which will occupy a large amount of class time. However, teachers cannot be online all day long, and it is difficult to give timely and rapid replies to students after class. When teachers fail to pay attention to students' online questions, students get bored and reduce their learning efficiency. When students have anxious academic emotions, they cannot master the knowledge they have learned well, and their thinking efficiency is not high. Some negative emotions will damage learners' efforts and affect learners' use of efficient learning strategies. After the end of the online class, the discussion area can be set up to reply to students' messages.

In the learning process, teachers can provide practical problems for students through the online platform, guide students to use online learning resources to sort out knowledge points, and create a learning atmosphere for students to study independently and actively communicate, which can reduce students' anxiety and effectively improve their learning efficiency. Students' learning attitude helps to stimulate their motivation, develop their potential and adjust their learning, among which students' emotional state in learning will have an impact on students' learning ability. Online learning for middle school students is full of intense emotional experience; the interaction among students and teachers will have some emotional experience. When students' positive emotions are enhanced, they will continue to study online. When students have anxiety during study, it will affect their learning continuity and hinder their learning efficiency. Teachers can make use of the rich online resources to design courses, give students the opportunity to discuss freely, flexibly adjust the class atmosphere, help students generate more positive emotions, cultivate students' good learning attitude, and guide students to study efficiently.

5. Conclusion

Contemporary college students are faced with the network environment supported by the Internet and mediated by information resources. On the one hand, with the continuous development and change of computer technology and network technology, the impact of network environment on college students' mental health is multifaceted and all-round. On the other hand, with the improvement of material life, people pay more attention to students' mental health. Therefore, this study takes college students' mental health as the research object to study the close relationship between students' emotion and learning in the network environment. In the context of modern educational technology, teachers can make full use of Internet resources and select examples combining current hot phenomena when explaining knowledge points to attract students' interest in learning and let students explore knowledge points freely. Modern educational technology model puts forward higher requirements for teachers. Teachers should not only learn to use network technology but also guide students to have a positive emotional state of learning. In the network classroom, teachers can obtain a variety of teaching methods

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9795676, 1 page
https://doi.org/10.1155/2023/9795676

Retraction

Retracted: Empirical Analysis of Enterprise Financial Management Risk Prediction in View of Associative Memory Neural Network

Security and Communication Networks

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Cheng and X. Zhang, "Empirical Analysis of Enterprise Financial Management Risk Prediction in View of Associative Memory Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 7825000, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7825000, 12 pages
https://doi.org/10.1155/2022/7825000

Research Article

Empirical Analysis of Enterprise Financial Management Risk Prediction in View of Associative Memory Neural Network

Huifang Cheng[1] and Xishuan Zhang[2]

[1]*School of Finance and Economics, Henan Polytechnic University, Jiaozuo, Henan 454003, China*
[2]*Research Center for Energy Economics, School of Business Administration, Henan Polytechnic University, Jiaozuo, Henan 454003, China*

Correspondence should be addressed to Xishuan Zhang; zhangxishuan@hpu.edu.cn

Received 19 February 2022; Revised 30 March 2022; Accepted 5 April 2022; Published 11 May 2022

Academic Editor: Chin-Ling Chen

Enterprise financial analysis has a far-reaching impact on modern enterprise management decision-making and plays a role that cannot be ignored. Financial status is related to the life and death of an enterprise and is the lifeline of an enterprise. Fast and efficient financial analysis can provide reliable and accurate decision-making information support for enterprise investors, operators, creditors, and other organizations and individuals to understand and evaluate the enterprise status and future development potential. With the development of intelligent methods such as associative memory neural network, the research of financial analysis decision support based on artificial intelligence has been paid more and more attention by academia and management, and has made new progress. Efficient and accurate financial risk prediction can help enterprises predict the possible financial risks in the future earlier, facilitate the early detection of problems, and take effective measures to avoid risks or minimize losses. However, most of the existing mature financial risk prediction studies are based on balanced data sets. The research on the classification of unbalanced data sets is not mature and perfect and needs to be further studied. Generally speaking, this paper mainly adopts the research method of cross integration of various disciplines and organically integrates the key theories, methods, and technologies such as default risk management theory, financial index analysis theory, data mining principle, prediction and decision theory, computer technology, multiclassifier integration technology, a variety of enterprise financial risk early warning technology and statistical sampling, carrying out systematic research on enterprise financial risk prediction. This review constructs the enterprise financial risk prediction method system based on heterogeneous data mining technology, mainly including data preprocessing layer, improved nearest neighbor delta increment layer, heterogeneous nearest neighbor extraction layer, and case-based reasoning prediction layer, so as to improve the traditional financial risk prediction method and obtain a new risk classification prediction model. The case-based method has some significant advantages in risk prediction performance, and it also helps to reduce the probability of financial risk.

1. Introduction

With the rapid development of China's economy, the financial capital market has become more and more competitive, and the demand for enterprise financial risk analysis and prediction has become more and more urgent. The majority of managers and investors urgently hope to establish an effective financial risk intelligent analysis and prediction system through continuously improving the prediction mechanism. Analyze and predict the overall financial risk of the enterprise, intelligently. Many companies have been implicated in the world economic crisis, their operations have been frustrated, consumer demand and investor confidence have declined, and they are facing the business difficulties of capital chain rupture, financial deterioration, difficult operation, and even helpless bankruptcy, which affect the stability and development of the global economy. How to predict the financial distress, prevent the further deterioration of the enterprise's financial situation, restructure the company's resources and structure, improve the comprehensive operation level, enhance the confidence of consumers and investors, and help the enterprise avoid or get rid of the known or unknown risks [1–10].

After years of practice, early excellent enterprises usually increase their profits by reducing costs and improving the durability of goods. However, this change has been unable to meet the competition mode under the new economic normal. Innovation has long been the secret of development and prosperity. It is undoubtedly the symbol of a new era. Therefore, in recent years, the research on quality and cost has been insufficient to meet the needs of some advanced enterprises. In order to surpass and distance themselves from other enterprises and compete for business opportunities, these enterprises focus on innovation and reducing their financial risks. Therefore, innovative enterprises were born. If innovation is separated from the enterprise, all the development strategies and projects about innovation will exist inefficiently. Shortening the innovation cycle and enhancing risk awareness are the key factors to enhance the competitiveness of enterprises. Compared with other types of enterprises, innovative enterprises have the characteristics of high risk, uncertainty, and high return. Therefore, compared with the financial risk management of ordinary enterprises, the financial risk management of innovative enterprises has both similarities and differences [11–15].

Financial risk classification and prediction are an important part of risk management and also an effective means to avoid or reduce risks. Financial risk classification and prediction refers to the establishment of financial risk classification and identification system according to the risk factors of historical financial data, which provide some valuable reference opinions and measures to support the decision-making of senior management of enterprises. At present, the research on enterprise financial risk classification and prediction can be roughly divided into two categories. One is the research on qualitative financial risk classification and prediction, which mainly uses subjective and artificial evaluation systems such as decision-making evaluation, and brainstorming to judge and analyze new things and problems through the summary of past individual or group experience, which often has certain limitations. One is quantitative financial risk classification and prediction research, which mainly adopts quantitative evaluation systems such as mathematics and computer data mining. Based on the analysis and demonstration of a large number of data, it often lacks certain flexibility and needs to be further improved or improved by constructing corresponding mathematical analysis models to make decision analysis on new things. Based on various related theories and technologies such as financial risk management theory, financial index analysis theory, data mining principle, prediction and decision theory, computer technology, multimodel fusion technology, multiclassifier recognition and analysis integration technology, data statistical analysis technology, case-based reasoning, and its reuse technology, this paper will study the difficult problems of enterprise financial risk prediction and evaluation; so as to overcome the current research deficiencies in this field and help to further expand the enterprise financial risk assessment modeling theory and method system, which has important theoretical research value [16–22].

Risk refers to the combination of the possibility and consequences of a specific adverse situation. In enterprise management, risk is divided into operational risk and financial risk. The financial risk refers to the possibility that the actual financial results of the enterprise deviate from the expected financial results and suffer losses in a certain period of time due to the internal and external environment and various unpredictable or uncontrollable factors in the process of various financial activities in the production and operation of the enterprise. At present, there are two main understandings of financial risk in academic circles: the first way of understanding holds that financial risk only exists in the process of enterprise financing, which is a narrow way of understanding. Under this way of understanding, the definition of financial risk refers to the risk caused by the failure of enterprises to repay their debts due to maturity. The other is a broad understanding, which holds that the financial risk of enterprises does not only exist in the financing process, but will exist as long as there is the flow of funds. Therefore, the financial risk in a broad sense is caused by the inconsistency between the actual income and the expected income in the whole process of capital operation due to a variety of unpredictable reasons.

The operation process of any enterprise's capital should include financing, investment, capital recovery, and capital distribution. There will be financial risks in each link of the capital operation process. The first is the financing link. When the enterprise raises funds in the early stage, due to the wrong selection of financing quantity or financing method, it has no way and ability to repay the loan after maturity, and the risk will exist in the financing link. The second is the investment link. When an enterprise invests, if it fails to make a feasibility analysis or makes mistakes in investment decision before investment, its actual rate of return will easily deviate from the expected return after maturity. This uncertainty of investment return is investment risk. The third is the recovery of funds. When the cash outflow generated by the enterprise in the process of operation cannot be recovered on time or directly, there will be a risk of fund recovery. Finally, the capital distribution link. When the enterprise makes profits after paying taxes at the end of the year, the capital distribution is unreasonable, which cannot meet the expected requirements of shareholders or ensure the development needs of the enterprise. At this time, the risk will inevitably exist in the capital distribution link. Therefore, financial risk exists in every link in the process of enterprise capital flow. As long as there is a risk in one link, it is likely to cause financial crisis [23–34].

As a branch of risk management, financial risk management, as an exante management function, is a management science developed on the basis of predecessors' rich risk management experience and modern scientific and technological achievements. Financial risk management means that enterprises or other business entities analyze and study various possible risks in advance in their respective business processes, preset corresponding countermeasures according to various risks, prevent and control them, solve them with economic and reasonable methods, and ensure the normal development of business activities, management process to ensure that their own economic interests are protected from loss. Financial risk management is mainly composed of risk identification, measurement, prevention, and control, the core of which is to effectively identify and measure all kinds of risks encountered. All these works are done to reduce financial risks and reduce enterprise losses as much as possible. Therefore, when making decisions on financial risk management, deal with the relationship between cost and efficiency, first consider solving the risk from the most economic and reasonable point of view, and formulate financial risk response strategies. The dynamic characteristics of risk determine that the management of financial risk is always a dynamic process. Due to the changing internal and external environment faced by the enterprise, in the implementation of financial risk management, the management plan to deal with financial risk should be adjusted in time in combination with the real-time change of financial risk status, and the behavior deviating from financial risk management should be continuously corrected. The following is a brief description of each link of financial risk management.

The rise of innovative enterprises is not accidental, but thanks to its strong technical strength. The rapid development of innovative enterprises has promoted the economic development of enterprises, China, and even the world. Its advantages and disadvantages coexist. Compared with traditional enterprises, innovative enterprises will encounter more risks. Because its main work is innovation, it will encounter many risks in the process of innovation. In order to eliminate the factors affecting the development of enterprises, to understand the risks from the root, and prevent and deal with various risks in time to prevent the occurrence of crises.

Enterprise financial risk management mainly relies on cutting-edge technologies such as computer and data mining to carry out real-time and systematic tracking and feedback on the internal financial operation status of the enterprise, monitor the possible financial risks, so that the enterprise can quickly and accurately respond to various financial risks, so as to avoid unnecessary or irreparable losses, and promote the enterprise to change the organizational structure, further promote organizational innovation and development. For any enterprise, financial problems are related to the life and death of the enterprise and are of great concern to the enterprise. Effectively predicting and avoiding enterprise financial risks is one of the important indicators to evaluate the operation and management level of an enterprise. The survival of enterprises, especially listed companies, directly or indirectly affects the interests of related parties. Strengthening the management of enterprise financial risk has become a common concern of contemporary business and academic circles. According to the theory of problem case similarity knowledge, this paper attempts to effectively predict enterprise financial risk through multimodel case-based reasoning reuse technology and its new fusion method.

By studying various factors that may make innovative enterprises fall into risk, hope to give enterprise managers some reference basis and help them formulate convenient, practical, and correct financial risk management methods. This paper uses specific innovative enterprises to carry out research. From the perspective of enterprise finance, this paper makes an in-depth exploration on the financial risks of innovative enterprises, studies and analyzes the occurrence mechanism, occurrence and future prevention measures of financial risks in innovative enterprises from the perspectives of the causes, conditions, and future management methods of financial risks. The global exponential stability of neural networks is analyzed and studied, and the effects of time delay and random disturbance on the stability of neural networks are fully considered. The global exponential stability of continuous time neural networks with constant time delays and their corresponding discrete models is studied. The conditions for the global exponential stability of the studied system and the necessary assumptions are given, and the vertex Lyapunov functional for each node is constructed. Based on the properties of equilibrium graph and topological property of network, the system is proved to be globally exponentially stable.

2. Associative Memory Neural Network

A bidirectional associative memory neural network model is established, and the nonlinear neuron excitation function in the model is neither bounded nor smooth. A sufficient condition for the exponential input-state stability of the bidirectional associative memory neural network model is obtained by using Lyapunov functional and linear matrix inequality [35–38]. The stability of associative memory neural network is obtained by Lyapunov functional and linear matrix inequality. In the later stage, the state stability of neural network with pulse and random will be studied by constructing new differential inequality methods.

The following bidirectional associative memory neural networks were studied:

$$\begin{aligned} \frac{\mathrm{d}x_i(t)}{\mathrm{d}t} &= -a_i x_i(t) + \sum_{j=1}^{n} u_{ij} f_j\left(y_j\left(t-\tau_j\right)\right) + s_i(t), i = 1, 2, \cdots, n, \\ \frac{\mathrm{d}y_j(t)}{\mathrm{d}t} &= -b_j y_j(t) + \sum_{j=1}^{n} u_{ii} g_j\left(x_i\left(t-\sigma_i\right)\right) + h_j(t), j = 1, 2, \cdots, n, \end{aligned} \tag{1}$$

where $x_i(t)$, $y_j(t)$ are the form states of the divine element, f_j, g_j are continuous, τ_j, σ_i are positive constants, a_i, b_j are nonnegative constants, and u_{ij}, s_i, u_{ii}, h_j are constants.

Each layer of the neural network model contains L neurons. Each neuron in the first layer is connected to the neuron in the second layer, and signals are transmitted bidirectional between the two layers. It is derived from a simple neural network in which neurons do not signal each other. Discrete time neural networks with constant time delays have a wide application background. Up to now, there are few research achievements. In recent years, the research in this field has been paid more and more attention. In this part, the discrete model corresponding to the continuous time neural network model with constant delay will be

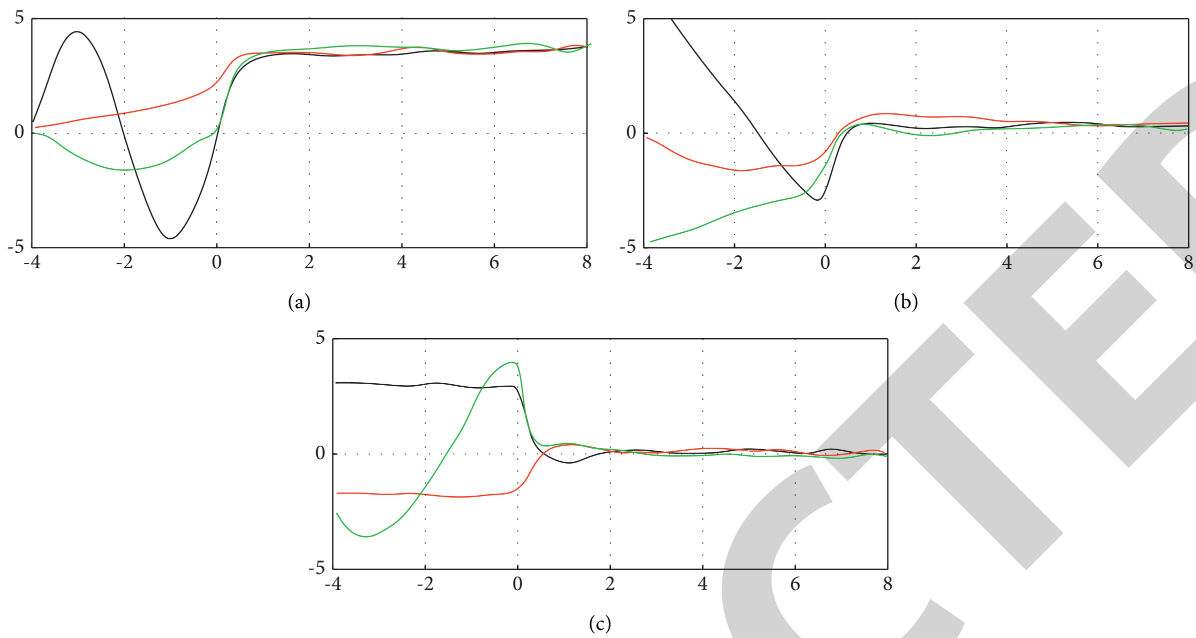


Figure 1: Equilibrium index of neural network.

established. There are many discretization methods for continuous time model. In this paper, a semidiscretization technique is used to obtain the corresponding discrete model of the system through two approximations, and it is proved that this discretization method can keep the stability of the original equilibrium point, which has strong practical significance. The sufficient criterion to ensure the global exponential stability of continuous time neural networks with constant time delays is established based on Lyapunov method and graph theory knowledge. Through practical application in real life, its stability has been effectively tested.

The output produced by neurons in one domain is immediately transmitted to neurons in the other domain as its input signal. This signal processing method may be consistent with the actual situation when the number of neurons is small, but if the number of neurons in the network is very large and the axon length of neurons is long, the signal will be delayed when it is transmitted from one neuron to other neurons.

$$\begin{aligned}\frac{\mathrm{d}x(t)}{\mathrm{d}t} &= -Ax(t) + Wf(y(t-\tau)) + s(t),\\ \frac{\mathrm{d}y(t)}{\mathrm{d}t} &= -by(t) + Vg(x(t-\sigma)) + h(t),\end{aligned} \tag{2}$$

where $x(t)$ and $y(t)$ are the form states of the divine element. A, W, b, and V are the Weight matrices.

The time delay represented by the neural network model is a constant time delay, that is to say, the state of the neuron at any time f is only related to the state before a fixed time length, which is obviously inconsistent with the actual situation.

The conventional methods to discretize a continuous system include Euler and Runge. Runge–Kutta method, etc. However, the dynamic properties of the discretized system may be quite different from those of its corresponding continuous system, which may lead to some conclusions obtained in the study of continuous dynamic system, such as the existence of equilibrium point and the criterion of stability cannot be used. The phenomenon that an exponentially stable equilibrium point of continuous neural network is no longer exponentially stable after being discretized by standard Euler method. If this happens, it becomes meaningless to use discrete systems to simulate, experiment, or calculate continuous systems. Therefore, to discretize a continuous system, it must be required that the discrete system inherit the dynamic properties of its corresponding continuous system in Figure 1. From the figure, we know that the dynamic characteristic refers to the relationship between the output and input of the system when the input of the detection system is a signal that changes with time. The main performance indexes of dynamic characteristic include unit step response performance indexes in time domain and frequency characteristic performance indexes in frequency domain.

The output information of the neural network is in the form of grasping, that is to say, these periodic solutions of the joint complaint memory are required to be in stable state. Associative memory can be realized only when the neural network can stabilize on the periodic solution of associative memory. If the periodic solution of the neural network associated with memory is stable, then the neural network can eliminate these noises or make the fuzzy information clear, and finally associate with the stable periodic solution. Otherwise, the network output information may not be the information we need to associate. Stability is one of the core problems of neural network theory and application. It is the theoretical basis of associative memory and also provides a reliable guarantee for practical application. Therefore, the study on the existence and stability of periodic solutions of neural networks is of great guiding significance to researchers applying MAM neural networks in designing neural network systems:

$$L_i(t) = U_i(t) + \sum_{j=1}^{l} \left|b_{ij}\right| L_j e^{a\sigma_{ij}} \int_{t-\sigma_{ij}}^{t} V_j(s)\mathrm{d}s + V_i(t) + \sum_{j=1}^{l} \left|d_{ij}\right| M_j e^{a\sigma_{ij}} \int_{t-\tau_{ij}}^{t} U_j(s)\mathrm{d}s. \tag{3}$$

The above neural network is approximated twice, and the following discrete neural network model is obtained:

$$\begin{aligned} U_i^{(n+1)} &= U_i^{(n)} e^{-a_i h} + \theta_i(h) \sum_{j=1}^{l} b_{ij} f_j\left(V_j^{\left(n-w^{ij}\right)}\right) + \theta_i(h) I_i, \\ V_i^{(n+1)} &= V_i^{(n)} e^{-c_i h} + \varphi_i(h) \sum_{j=1}^{l} d_{ij} g_j\left(u_j^{\left(n-k^{ij}\right)}\right) + \varphi_i(h) J_i, \\ \theta_i(h) &= \frac{\left(1-e^{-a_i h}\right)}{a_i}, \\ \varphi_i(h) &= \frac{\left(1-e^{-c_i h}\right)}{c_i}. \end{aligned} \tag{4}$$

The vertex Lyapunov functional of the i-th node of the discrete-time neural network is constructed as follows:

$$H_i^{(n)} = U_i^{(n)} + \sum_{j=1}^{l} \left|b_{ij}\right| L_j \varphi_j(h) \lambda^{w_{ij}+1} \sum_{m=n-w_{ij}}^{n-1} V_j(m) + V_i(n) + \sum_{j=1}^{l} \left|d_{ij}\right| M_j \theta_j(h) \lambda^{k_{ij}+1} \sum_{m=n-ki_j}^{n-1} U_j^{(m)}. \tag{5}$$

Under the same constraints, it is proved that the discrete-time neural network is globally exponentially stable and has the same equilibrium point as the continuous time neural network, that is, the discrete-time neural network obtained by this discrete-time method inherits the stability state of the continuous time neural network. The global moment exponential stability of stochastic neural networks with time-varying delays is studied. The conditions for the exponential stability of the global moment of the studied system are given. Combining various inequalities, the neural network is established on the strongly connected graph, and the appropriate vertex Lyapunov functional is constructed. It is proved that the system is exponentially stable:

$$\begin{aligned} \mathrm{d}x_k(t) &= \left[-a_k x_k(t) + \sum_{h=1}^{l} b_{kh} f_h\left(y_h(t)\right) + \sum_{h=1}^{l} e_{kh} f_h\left(y_h\left(t-\tau_{kh}(t)\right)\right)\right] dt + \sigma 1\left(x_k(t), y_k(t)\right) dw_1(t), \\ \mathrm{d}y_k(t) &= \left[-c_k y_k(t) + \sum_{h=1}^{l} d_{kh} g_h\left(x_h(t)\right) + \sum_{h=1}^{l} p_{kh} g_h\left(x_h\left(t-\sigma_{kh}(t)\right)\right)\right] dt + \sigma 2\left(x_k(t), y_k(t)\right) dw_2(t), \\ \frac{\mathrm{d}u_i(t)}{\mathrm{d}t} &= -a_i u_i(t) + \sum_{j=1}^{3} b_{ij} \operatorname{arc}\tan\left(v_j\left(t-\sigma_{ij}\right)\right) + I_i, \\ \frac{\mathrm{d}v_i(t)}{\mathrm{d}t} &= -c_i v_i(t) + \sum_{j=1}^{3} d_{ij} \sin\left(u_j\left(t-\tau_{ij}\right)\right) + J_i, \end{aligned} \tag{6}$$

a_i, c_i, $b_{i\,j}$, $d_{i\,j}$ are nerve parameters, I_i, J_i are external input values.

The general conditions for the p-order exponential stability of the neural network are given, and the applicability of the conditions is proved by combining the inequality.

3. Empirical Analysis of Enterprise Financial Management Risk Prediction

By integrating the basic category prediction results in a weighted way, the prediction results of the financial management system on this test sample can be obtained, and then the test accuracy of the group on this test sample can be obtained. This process is repeated until each sample has performed a training and testing process. The accuracy of prediction is verified by averaging the above accuracy.

As shown in Figures 2 and 3, for both N2 and N3, the left side of Figure 2 is the relationship between sample points and prediction accuracy, and the right side of Figure 2 is the relationship between sample points and personal optimal values. The prediction accuracy generally increases first and then decreases with the increase of the basic number, but it is not stable and monotonous, which proves the correctness of the forward search and backward pruning strategies. The prediction accuracy of the system is higher than or equal to the optimal single classifier, which proves the effectiveness of the proposed method. In practice we can according to the need, if you need the most concise combination system, all can choose a combination of 3, but if you want to get the highest prediction accuracy, can choose 7, which is composed of seven basic classifier combination system, so the method of this section makes it practical in the process of financial distress prediction obtained greater flexibility. To some extent, it proves the importance and effectiveness of

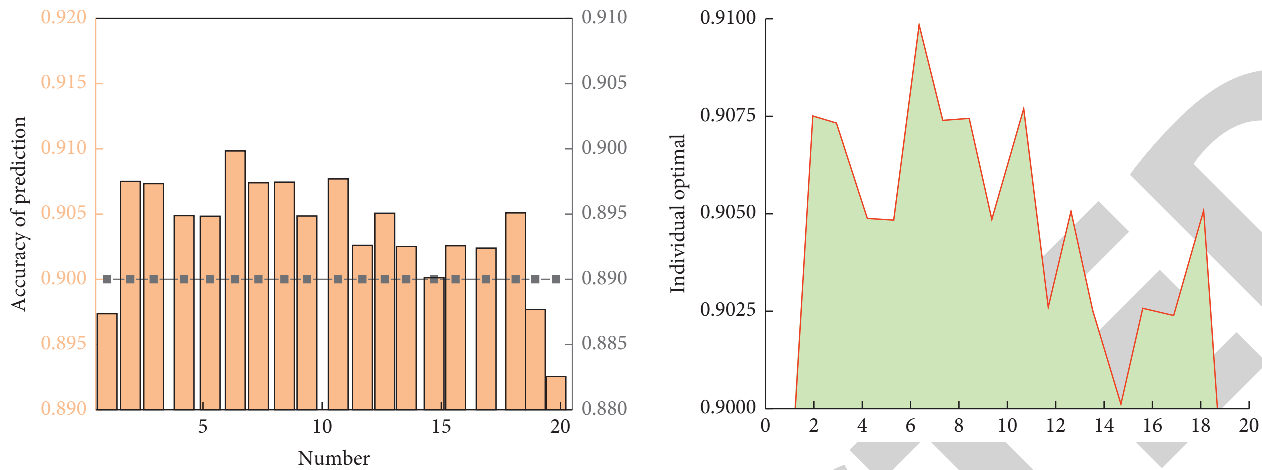


Figure 2: System prediction accuracy of N2.

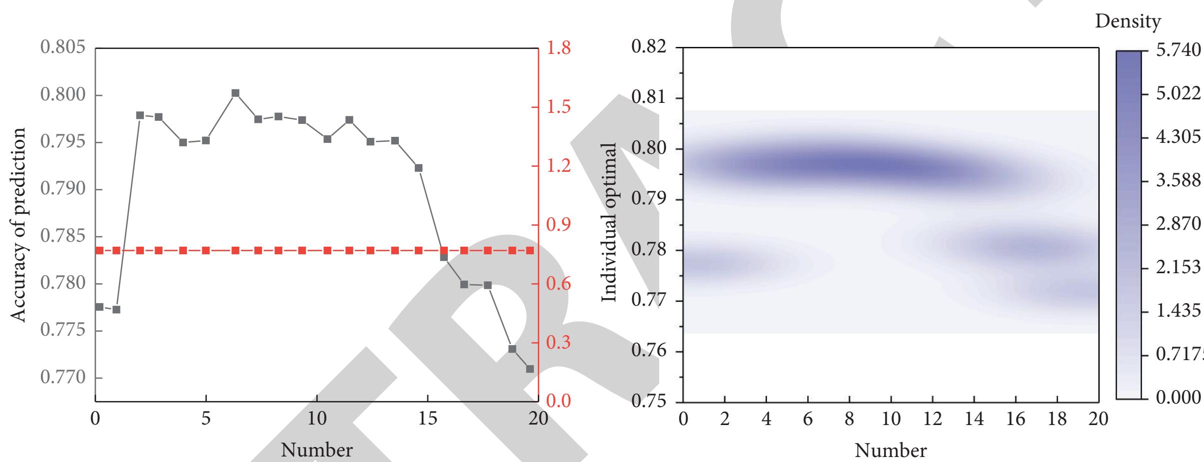


Figure 3: System prediction accuracy of N3.

feature selection in classifier construction. In many cases, it indicates how well a model is performing, but in some cases, the number of subclassifiers cannot be too high or too small. In particular, the number of subclassifiers should be appropriately selected in the integration algorithm to achieve the optimal classification accuracy.

In order to further analyze the advantages and disadvantages of the forward greedy search strategy proposed in this paper, it is compared with several other commonly used selective integration strategies. The experimental results are shown in Figures 4 and 5. Note that the left side of Figure 5 is the relationship between sample points and prediction accuracy, the right side of Figure 5 is the relationship between sample points and personal optimal values. Among them, ascending integration means that basic classifiers are sorted from low to high according to individual performance. Starting from the worst single classifier, each basic classifier is added successively until all of them are added. Meanwhile, the combined performance of the new system after each addition is recorded. On the contrary, the descending ensemble sorts basic classifiers according to individual performance from high to low, and adds each suboptimal single classifier in turn starting from the optimal single classifier. Random search is a process in which each single classifier in the basic classifier library is added into the system one by one by selecting one from the basic classifier library as the starting point and adopting the sampling method without putting back. The system consisting of forward greedy search occupies the highest position, followed by descending integration and random search, and ascending integration occupies the lowest position. Although the trend of ascending integration shows a trend of rapid rise, due to its low starting point, although its performance is greatly improved compared with the starting point, it can only reach or approximate the level of individual optimal single classifier, and even worse than the prediction performance of random combination. Therefore, if the prediction ability of a single classifier outside the system can be achieved, even if the prediction ability of the combined system has been improved to some extent compared with that inside the system, such improvement is meaningless, because it does not conform to the principle of cost-effectiveness. Therefore, it is meaningful to promote on the basis of the optimal individual, which proves the correctness of selecting the

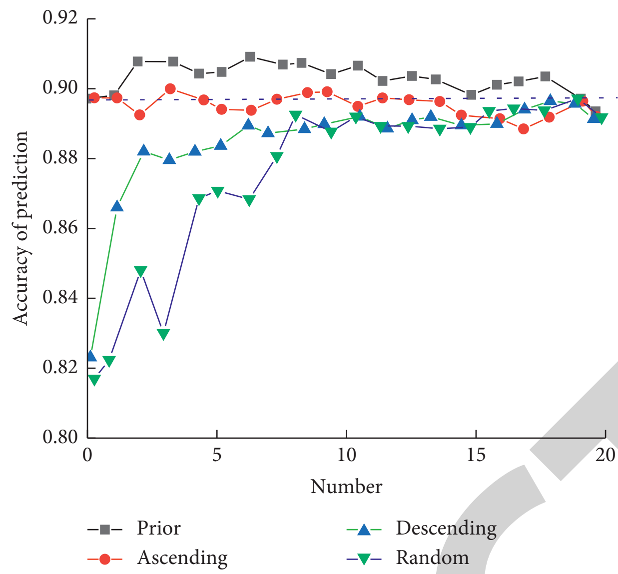


Figure 4: Performance comparison of combinations of different fusion sequences.

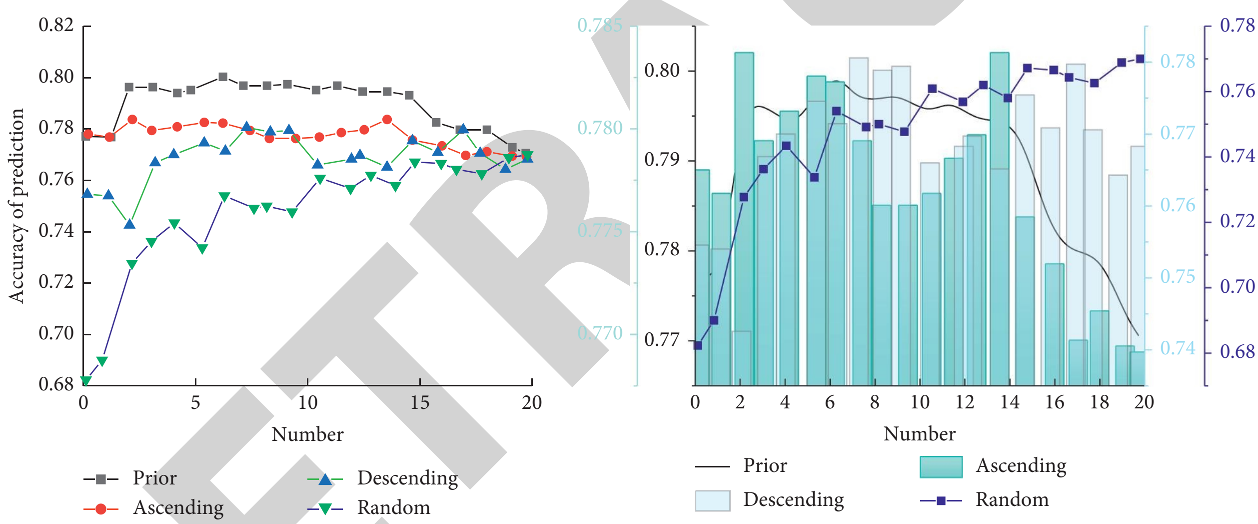


Figure 5: Different fusion sequences.

individual optimal basic classifier in this paper and constructing the combined system from it. For forward greedy search and descending integration, although both of them start from the individual optimal basic classifier, it can be seen that the performance of descending integration is still very unstable. The performance of the combined system vibrates up and down the individual optimal line, while the performance of forward search is significantly higher than the individual optimal line. The combined performance of the first few basic classifiers with better performance may not be better, or even worse, indicating the importance of differences between basic classifiers for system construction. In each step, the forward greedy search retains the best combination prediction performance in the current situation (given number of basic classifiers), and its work starts from the difference. Although it has not gone through large-scale system combination trial calculation, its simple optimization of the current "greedy" thought, or provides a large number of satisfactory solutions for the postpruning work.

For model in Figure 6, static model in the first three year of testing accuracy is higher than the fixed window model, may be due to a few year of annual training samples and testing samples are relatively close, concept drift phenomenon is not obvious, but with the passage of time, the sample of old phenomenon is more serious, and its testing accuracy drop is more obvious. However, the test accuracy of the fixed time window method is relatively stable over time, which further illustrates the importance of updating the model. The dynamic model with fixed time window is superior to the static model. Built in a certain period of time and, therefore, is not too up to date, and the static model often cannot meet the needs of financial distress prediction, even if the enterprise financial predicament definition is unchanged under the premise of characteristics of the concept of financial distress,

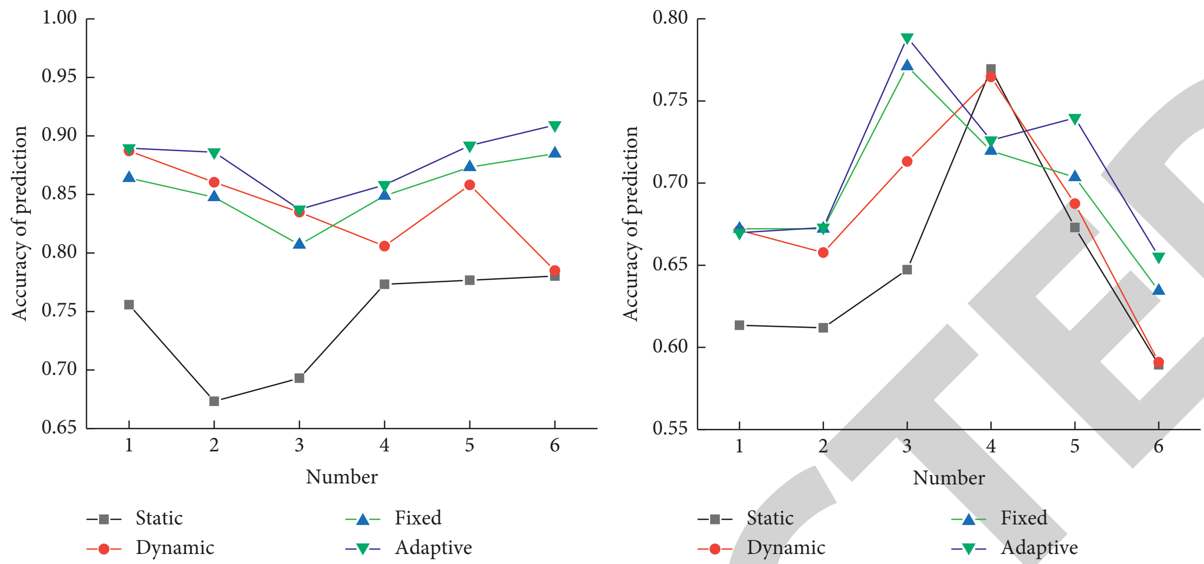


FIGURE 6: Line chart of under different characteristics.

and the changes are with the change of economic environment, and the financial difficulties of the virtual concept drift, but the enterprise is necessarily based on the current available economic information. Dynamically update the financial distress prediction model in time to continuously meet the needs of future financial distress prediction.

Dynamic selection and dynamic integration model accuracy is mostly higher than the window size adaptive method in Figure 7, although this is because the window size adaptive method can be found out with the current financial difficulties concept features and the most relevant data sample subset, but in financial distress prediction for future samples, build a single financial distress prediction model, The dynamics is reflected in the selective adjustment of modeling training samples. And dynamic selection and dynamic integration model is first of all, using the method of adaptive window size selection of the training sample library building basic classifier, completion of the current financial difficulties concept features of different information levels of digging, and according to the specific characteristics of the future to predict a single sample, dynamic selecting one or a few of the most relevant for decision-making. This dynamic is described for a single sample level, and the study is more detailed and a more scientific theory. Of the dynamic selection and dynamic integration methods, no one method is always optimal over the years advancing over time, and all methods have their advantages and relative limitations. It can be seen that simple combination integration prediction cannot effectively improve the accuracy of financial distress prediction facing concept drift, which proves the effectiveness and necessity of dynamic selection and dynamic integration from another perspective.

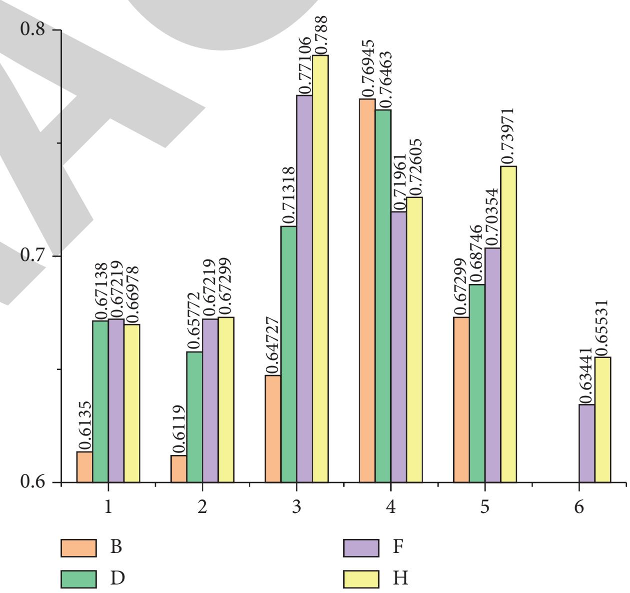


FIGURE 7: Accuracy rate.

4. Construction of Enterprise Financial Risk Management and Risk Control Mode

Establishing an innovative enterprise financial risk management model can sort out the financial risks in the process of enterprise development, formulate and implement prevention and control measures, especially strengthen the management of the enterprise's major strategy formulation, important domestic and foreign investment, and the company's operation and management objectives, so as to effectively prevent the occurrence of major financial risks. To make innovative enterprises, we must have plans and treatment measures for major and sudden financial risk events, so that they can deal with financial risks in an emergency, so as to prevent huge losses to the company. We must establish a financial risk management system that basically covers all businesses and management and carries out financial risk management around the company's strategic development objectives, business management objectives, and subobjectives, so as to make risk management an effective way for the company to enhance execution, improve comprehensive

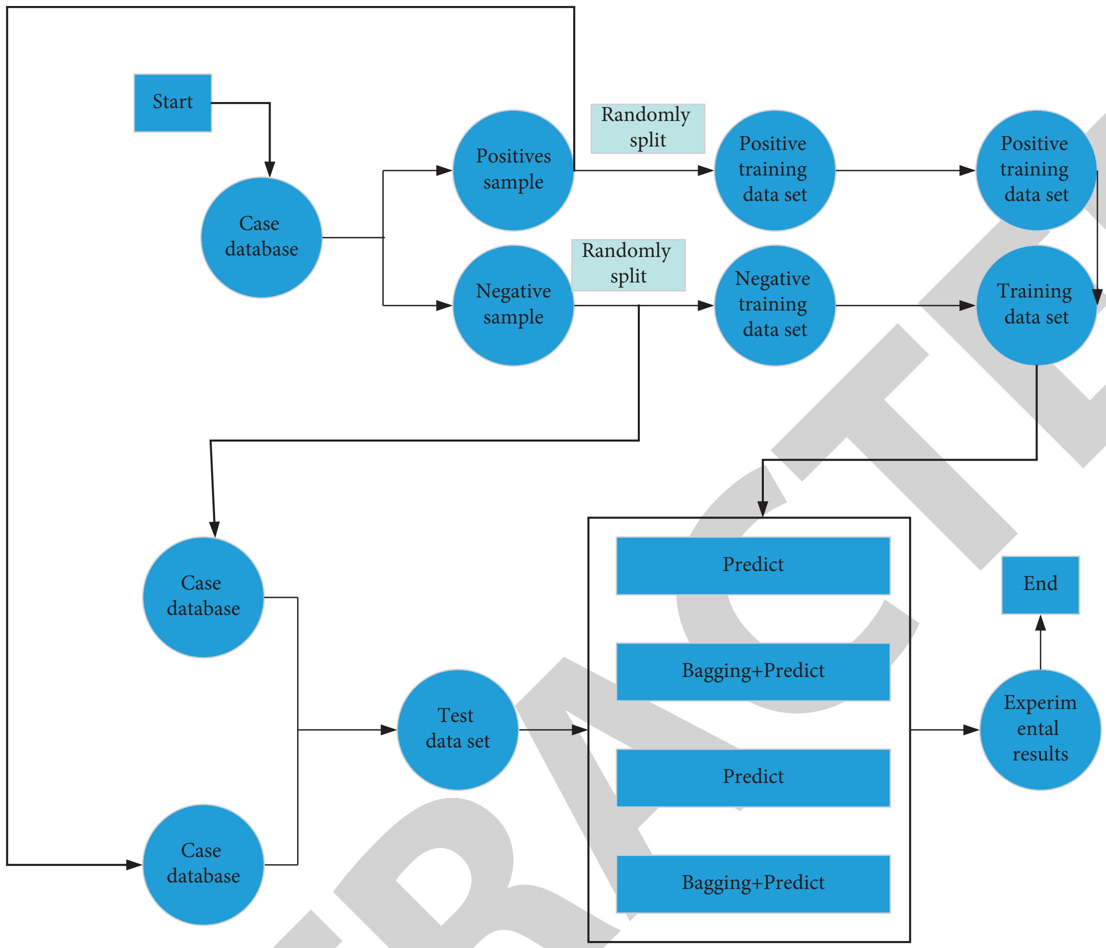


Figure 8: Predictive model test.

strength, improve operation quality, create value and make scientific decisions, and also continuously improve the company's operation and management level and profitability as in Figure 8.

The premise of building an innovative enterprise financial risk management model is to ensure the effectiveness of the model. Only if the model is effective, the follow-up work is meaningful. Before the construction of this model, we must have an in-depth understanding of the characteristics of enterprises and departments to ensure that the designed management framework is in line with the studied enterprises and can enable enterprises to successfully avoid risks. At the same time, the system involved in the management mode should also conform to the characteristics of the enterprise itself, rather than a simple system design for all enterprises as a whole. Effectiveness is the core feature of risk management model. With more and more innovative enterprises in China, there will be more and more types of management modes, which requires R&D personnel to devote more efforts to the research of target enterprises. The construction mode should fully consider the relevant factors that need to be paid attention to as much as possible.

The financial risk of innovative enterprises is often inseparable from personnel. Innovative enterprises mainly rely on innovation. Only be people will be innovative. Products depend on people to design, and follow-up work such as production and sales also need people to complete. Enterprises at the cultural level need to rely on people to formulate overall strategies and set overall goals. However, due to different opportunities, different ideas, different levels of education, and different knowledge reserves for R&D, the awareness of financial risk of innovative enterprises is also different. Therefore, it is necessary to improve the risk awareness of R&D personnel when building an innovative financial risk management framework. At the same time, the construction of risk management framework needs to pay attention to human nature and the staff unite as one to make the model achieve the best effect.

When constructing the financial risk management model for enterprises, the framework is appropriately adjusted in combination with the characteristics of innovative enterprises, and some factors such as its two dimensions are extended. The implementation of financial risk management is promoted according to the idea of "overall planning and step-by-step implementation" in accordance with the reality, seeking practical results, combined with its own business positioning and operating characteristics. And continuously update the financial risk management with the development of the company and the continuous improvement of operation and management requirements.

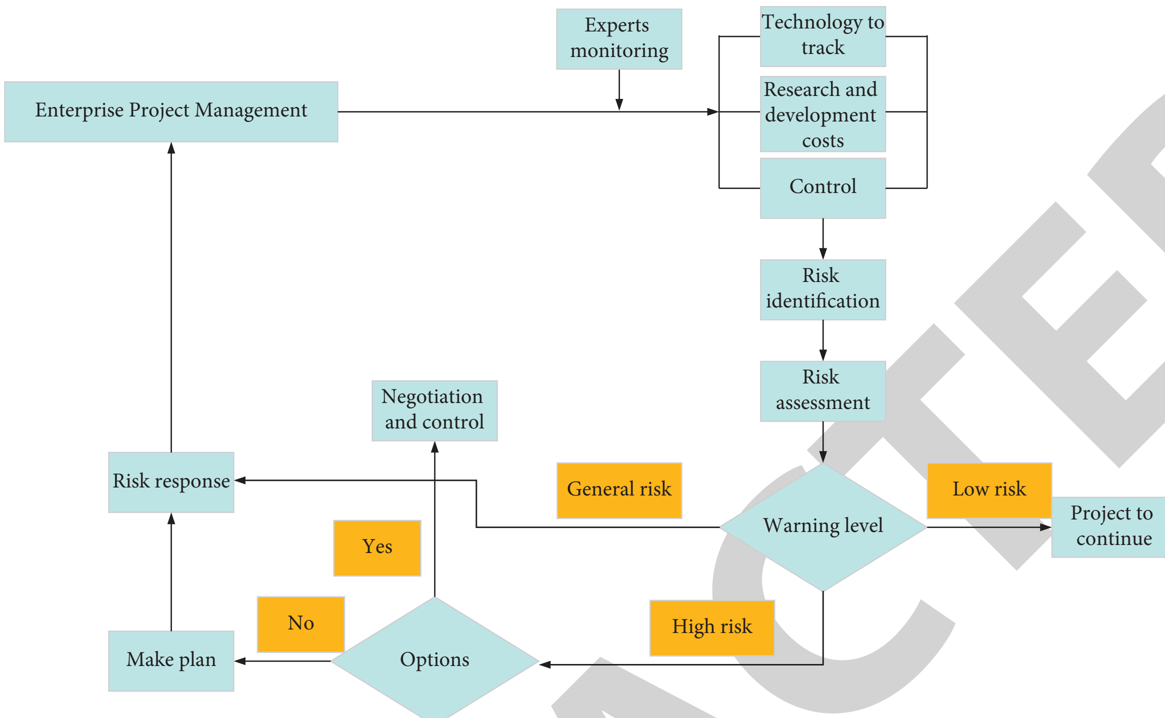


Figure 9: Flow chart of risk control.

After the enterprise passes the risk identification, the next task is to evaluate the output risk index results. The output results of the risk management process can be divided into three levels. The first is low risk, which can be appropriately ignored, and the risk management mechanism does not need to give an alarm. The second is the general risk alarm. In this case, enterprises will receive the alarm to remind them that they should take necessary measures to resist the occurrence of risks. The third is high risk. Issuing a high risk alarm indicates that the enterprise is already quite hazardous and needs to solve the risk as soon as possible. At this time, the committee needs to check the existing scheme library to see whether there is a scheme to deal with the risk. If not, it needs to carefully study the risk according to the risk situation and formulate effective schemes to deal with the risk in time, At the same time, it is stored in the scheme Library in case of being caught off guard in similar situations next time. The following is the risk handling flow chart set by the enterprise project risk management committee in Figure 9.

5. Conclusion

(1) The analysis framework breaks the financial distress modeling idea of giving a static model structure in advance on the static data set and emphasizes the role of data, that is, the classification and prediction knowledge of financial distress. The flow sample model further breaks the limitation of the static expansion period and ensures the prediction effectiveness of the financial distress prediction model over time through the dynamic update of the model, which is more scientific and meets the needs of practical application. In terms of static data, taking the system prediction accuracy as the optimization goal, two dynamic selective integration methods of multiclassifier systems are proposed from the perspective of local optimization and global optimization, which breaks the static limitation of previous combined systems.

(2) The experimental results show that based on the greedy search method designed at each step and the overall pruning strategy, the prediction performance of the combined system obtained by dynamic mining is significantly better than that of the static fully integrated system and the individual optimal single classifier. The effectiveness of this method is further illustrated in the subsequent comparative study of various search sequences. A dynamic selective integration financial distress prediction method based on genetic search is designed. The experimental results show that without considering the time cost, the performance of the optimal prediction system obtained by this global search selective integration method is better through large-scale optimization trial calculation in the feasible solution space.

(3) Financial risk exists objectively and plays an important role in the profit and loss of the enterprise and how it operates. It is unrealistic to completely eliminate the risk and its impact. This objective existence determines that we fully understand and avoid it in the investment process. Financial risk runs through the enterprise's financial system and is reflected in various financial relations. The financial system is affected by many uncertain factors.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Henan Province Key Scientific Research Project Program Support, "Research on Constructing a Monitoring Mode of Online Teaching based on Big Data" (Grant no. 22A880009).

References

[1] W. M. Badenhorst, "Fair value intensity and analyst forecasts," *Journal of Economic and Financial Studies*, vol. 11, no. 1, pp. 1–12, 2018.

[2] G. E. Batta, J. Qiu, and F. Yu, "Credit derivatives and analyst behavior," *The Accounting Review*, vol. 91, no. 5, pp. 1315–1343, 2016.

[3] M. Bradshaw, Y. Ertimur, and P. O'Brien, "Financial analysts and their contribution to well-functioning capital markets," *Foundations and Trends in Accounting*, vol. 11, no. 3, pp. 119–191, 2017.

[4] D. Breeden and S. Viswanathan, "Why do firms hedge? An asymmetric information model," *Journal of Fixed Income*, vol. 25, no. 3, pp. 7–25, 2016.

[5] B. Gumb, P. Dupuy, C. R. Baker, and V. Blum, "The impact of accounting standards on hedging decisions," *Accounting, Auditing & Accountability Journal*, vol. 31, no. 1, pp. 193–213, 2018.

[6] M. L. Sang and S. R. Jin, "Ambidextrous supply chain as ADynamic capability building a resilient supply chain," *Management Decision*, vol. 54, no. 1, pp. 2–23, 2016.

[7] S. Duan, X. Hu, and Z. Dong, "Memristor-based cellular nonlinear/neural network: design, analysis, and applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 6, pp. 1202–1213, 2017.

[8] H. Wang, S. Duan, and T. Huang, "Exponential stability of complex-valued memristive recurrent neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 766–771, 2016.

[9] S. Duan, H. Wang, and L. Wang, "Impulsive effects and stability analysis on memristive neural networks with variable delays," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 2, pp. 1–6, 2016.

[10] T. Chen, L. Wang, and S. Duan, "Implementation of circuit for reconfigurable memristive chaotic neural network and its application in associative memory," *Neurocomputing*, vol. 380, no. 2, pp. 36–42, 2020.

[11] X. Zeng, Z. Xiong, and C. Wang, "Hopf bifurcation for neutral-type neural network model with two delays," *Applied Mathematics and Computation*, vol. 282, no. 13, pp. 17–31, 2016.

[12] Y. Sheng and Z. Zeng, "Impulsive synchronization of stochastic reaction-diffusion neural networks with mixed time delays," *Neural Networks*, vol. 103, no. 5, pp. 83–93, 2018.

[13] X. Yang, Q. Song, J. Cao, and J. Lu, "Synchronization of coupled Markovian reaction-diffusion neural networks with proportional delays via quantized control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 951–958, 2019.

[14] W. He, T. Luo, and Y. Tang, "Secure communication based on quantized synchronization of chaotic neural networks under an event-triggered strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3334–3345, 2019.

[15] H. Achouri and C. Aouiti, "Homoclinic and heteroclinic motions of delayed inertial neural networks," *Neural Computing & Applications*, vol. 33, no. 12, pp. 6983–6998, 2021.

[16] M. S. Ali, S. Saravanan, M. E. Rani et al., "Asymptotic stability of cohen-grossberg BAM neutral type neural networks with distributed time varying delays," *Neural Processing Letters*, vol. 46, no. 3, pp. 991–1007, 2017.

[17] J. Bai, R. Lu, A. Xue, Q. She, and Z. Shi, "Finite-time stability analysis of discrete-time fuzzy Hopfield neural network," *Neurocomputing*, vol. 159, pp. 263–267, 2015.

[18] Y. Liu, C. Li, T. Huang, and X. Wang, "Robust adaptive lag synchronization of uncertain fuzzy memristive neural networks with time-varying delays," *Neurocomputing*, vol. 190, pp. 188–196, 2016.

[19] P. Liu, Z. Zeng, and J. Wang, "Global synchronization of coupled fractional-order recurrent neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 8, pp. 2358–2368, 2019.

[20] C.-Y. Cheng, K.-H. Lin, C.-W. Shih, and J.-P. Tseng, "Multistability for delayed neural networks via sequential contracting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 12, pp. 3109–3122, 2015.

[21] J. Chen, J. H. Park, and S. Xu, "Stability analysis of discrete-time neural networks with an interval-like time-varying delay," *Neurocomputing*, vol. 329, pp. 248–254, 2019.

[22] J. Zhang, L. Ma, and Y. Liu, "Passivity analysis for discrete-time neural networks with mixed time-delays and randomly occurring quantization effects," *Neurocomputing*, vol. 216, pp. 657–665, 2016.

[23] Y. Shan, K. She, S. Zhong, Q. Zhong, K. Shi, and C. Zhao, "Exponential stability and extended dissipativity criteria for generalized discrete-time neural networks with additive time-varying delays," *Applied Mathematics and Computation*, vol. 333, pp. 145–168, 2018.

[24] A. Wu, Z. Zeng, and X. Song, "Global Mittag-Leffler stabilization of fractional-order bidirectional associative memory neural networks," *Neurocomputing*, vol. 177, pp. 489–496, 2016.

[25] Z. Ding, Y. Shen, and L. Wang, "Global Mittag-Leffler synchronization of fractional-order neural networks with discontinuous activations," *Neural Networks*, vol. 73, pp. 77–85, 2016.

[26] F. Wang, X. Liu, M. Tang, and L. Chen, "Further results on stability and synchronization of fractional-order Hopfield neural networks," *Neurocomputing*, vol. 346, pp. 12–19, 2019.

[27] T. Hu, Z. He, X. Zhang, and S. Zhong, "Global synchronization of time-invariant uncertainty fractional-order neural networks with time delay," *Neurocomputing*, vol. 339, pp. 45–58, 2019.

[28] T. Hu, X. Zhang, and S. Zhong, "Global asymptotic synchronization of nonidentical fractional-order neural networks," *Neurocomputing*, vol. 313, pp. 39–46, 2018.

[29] Y. Gu, H. Wang, and Y. Yu, "Stability and synchronization for Riemann-Liouville fractional-order time-delayed inertial

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9873758, 1 page
https://doi.org/10.1155/2023/9873758

Retraction

Retracted: Optimization of Printing and Dyeing Energy Consumption Based on Multimedia Machine Learning Algorithm

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Zhang and Y. Yu, "Optimization of Printing and Dyeing Energy Consumption Based on Multimedia Machine Learning Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 1960425, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1960425, 11 pages
https://doi.org/10.1155/2022/1960425

Research Article

Optimization of Printing and Dyeing Energy Consumption Based on Multimedia Machine Learning Algorithm

Xulan Zhang [1] **and Yue Yu** [2]

[1]China Knitting Industrial Association, Beijing 100020, China
[2]Dacheng County Power Supply Branch of State Grid Jibei Power Co., Ltd., Langfang 065900, China

Correspondence should be addressed to Yue Yu; 20162103826@mails.imnu.edu.cn

Received 15 February 2022; Accepted 18 March 2022; Published 10 May 2022

Academic Editor: Chin-Ling Chen

In order to solve the problem with the expansion of the industrial scale, the contradiction between energy resources and development of printing and dyeing enterprises must implement refined management and optimize the allocation of resources such as production technology, energy consumption types, and metering instruments. As a printing and dyeing textile industry with high energy consumption and high pollution, energy conservation and emission reduction have become difficult problem to be solved in this industry. The premise of optimizing resource allocation is to have an objective and scientific evaluation of the current energy resource allocation of printing and dyeing enterprises. In view of this, through the investigation of printing and dyeing enterprises, this paper puts forward the index system of enterprise energy consumption optimization evaluation. Based on the application of data warehouse and combined with historical data, a new energy consumption optimization evaluation method is proposed. This survey has basically understood the current situation of energy and water resources management of a printing and dyeing enterprise and pointed out the direction for the development of enterprise energy optimization project in the next step. By means of multimedia informatization, with the help of Internet of Things sensing technology, building a new generation of energy management system, improving the refined management of energy measurement, and improving the energy assessment system, enterprises can achieve significant economic and social benefits in terms of energy conservation and emission reduction. On this basis, a low-power scheduling strategy for typical data center applications is designed and implemented. The algorithm uses the input data of different calculation parts, performs matching to determine the redundant part of the calculation process, and schedules the algorithm. Experimental data show that the mean square error of the limit tree is 0.0004248 and the mean square error of the decision tree is 0.01581; through calculation, the algorithm can achieve 23% and 17% energy saving.

1. Introduction

With the expansion of data scale, how to use the machine learning algorithm to analyze these data in real time or off-line and find the internal law of data has become an important way for various industries to improve the accuracy of decision-making [1]. For example, Google indexes and ranks 571 new websites that emerge every minute based on machine learning algorithms; Wal-Mart, the world's largest retailer, needs to analyze and process more than 1 million transaction data per hour, so as to provide decision-making for its further business activities; Microsoft Research Asia, according to the air quality data provided by existing monitoring stations and other multiple data sources in the city, uses machine learning technology to fully analyze big data; it can be inferred in real time that urban air quality data contain fine particulate matter information [2, 3]. Therefore, machine learning has become an indispensable functional module in the data center. In this context, various machine learning systems running in data centers continue to emerge, such as GraphLab, Mahout, and MadLINQ. Among them, Mahout is a distributed machine learning library based on Hadoop, and it has been widely used by many companies (such as Yahoo, Twitter, and LinkedIn). In order to further optimize the consumption of printing and dyeing, we first conduct an in-depth characteristic analysis of typical machine learning algorithms for printing and dyeing [4].

Machine learning tasks are computationally intensive applications; therefore, its main energy consumption is reflected in data analysis and calculations. However, these algorithms are in the process of machine learning and need to continuously iteratively categorize and analyze data; there are a lot of redundant calculations in this process. These redundant calculations bring unnecessary energy consumption overhead. At the same time, in the data center, machine learning is mainly used for data analysis to find the internal laws of data, so there are no very strict requirements for the accuracy of users' calculation results. In the recommendation system, you only need to obtain the items that users like, and the overly accurate calculation results are of little significance here [5, 6]. With the rapid development of the Internet of Things and artificial intelligence, some industrial enterprises try to achieve better energy-saving and emission reduction effects by collecting the data generated in the production process and analyzing it with some sensor equipment to a certain degree. Collect some production data in the production process, such as energy consumption and water consumption, by installing relevant sensor equipment of relevant equipment in the production process. Using the corresponding machine learning algorithm in a large number of production data, we can find some inherent laws and patterns in the production data, so as to achieve more accurate energy conservation and emission reduction.

In terms of energy conservation and emission reduction, the initial exploration of enterprises is generally to optimize the industrial process to a certain extent and to upgrade some key equipment, so as to achieve energy conservation and emission reduction to a certain extent and promote the recycling of resources. In printing and dyeing enterprises, there are many different ways to achieve energy-saving effect, including process improvement, waste water and waste gas waste heat recovery, equipment update and maintenance, and the use of green lighting. In recent years, with the development of environmental protection trend, the printing and dyeing industry began to study the new technologies of energy saving, consumption reduction, and emission reduction, advocating the concept of green printing and dyeing, and was committed to transforming from extensive to intensive economic development form [7]. Based on the above characteristics of machine learning algorithms, we designed and implemented an energy-saving mechanism for machine learning calculations in data centers. It is proposed to remove redundancy by matching input data, so as to achieve energy-saving methods. The core idea is through twice input matching, the similarity of its output is analyzed, and the energy-saving effect is achieved by reusing the calculation results and reasonable scheduling [8, 9]. The experimental data shows that on the basis of ensuring the accuracy of the algorithm, the algorithm can effectively reduce the redundant calculation of the data center machine learning algorithm, so as to achieve the effect of energy saving. The machine learning algorithm is shown in Figure 1. Starting from the data of the printing and dyeing process finalization process, this paper combines order-related information, process parameter-related information, energy-related multimedia information, etc., to construct energy consumption categories and energy consumption as predicted values, and conduct model training after certain data preprocessing. Through the application of the optimization algorithm, the energy consumption in the subsequent production process is predicted, and the predicted data are used to adjust the process parameters to a certain extent so as to achieve the effect of energy saving and emission reduction [10].

2. Literature Review

Zhou and others said that with the new trend of social development, under the new wave of Internet+ and big data, the deep mining processing of data has entered saturation. This requires people to re-emphasize machine learning algorithms and set up a reasonable algorithm; let the machine do an in-depth analysis of the data by itself. This can be to a large extent solve the work that cannot be handled manually [11]. Xu and others expressed the application of machine learning algorithms; it is widely used in all aspects and various fields of today's society. For example, we use the Internet to search, use social platforms, and use e-commerce platforms [12]. Zhang and others obtained through data collection and used machine learning algorithms in the face of complicated data and information platforms on the Internet and the recommendation mechanism of shopping websites, which can analyze the user's points of interest and make better pushes [13]. Yin and others said that the application of machine learning algorithms in the abovementioned fields mainly reflected in the different data in the information that can be identified, for example, pictures, text, mixed graphics, audio, and so on; in addition, it can successfully convert voice to text output and identify the domain of the content of an article and the blocking of bad content [14]. Sosnicki stated that the advantages of this technology have been valued in many fields; from this, we can see the development prospects of this advantage [15]. Rao and others found through investigation that relatively speaking, the research of some machine learning algorithms focuses on the theoretical model of machine research. They attach great importance to the accuracy of the theoretical model and analyze various disadvantages faced after being put into use. In other words, in the field of machine learning algorithms, a number of mutually supporting algorithm linkage mechanisms have been introduced, and data classification creation mechanism, so with the help of multiangle models, carries out practical training to improve the feasibility of data processing tasks [16]. Su and others have found through research that the application of vector machines supported by machine learning algorithms is an important part of connecting other vector machines. Only by optimizing the scheme between vector machines can the optimal data classification for data processing be obtained. Based on the development of the era of big data, the speed of various information data processing has been accelerated. Based on the premise of development in the era of big data, various information data processing is speeded up [17]. Camejo and others said that from multiple perspectives of current social information data processing and traditional information data processing algorithms, some large-scale data are no longer applicable. In today's society where data flow and information flow are exploding, in-depth data mining has gradually developed into a new trend [18]. Khan and others have found through surveys that the transformation of

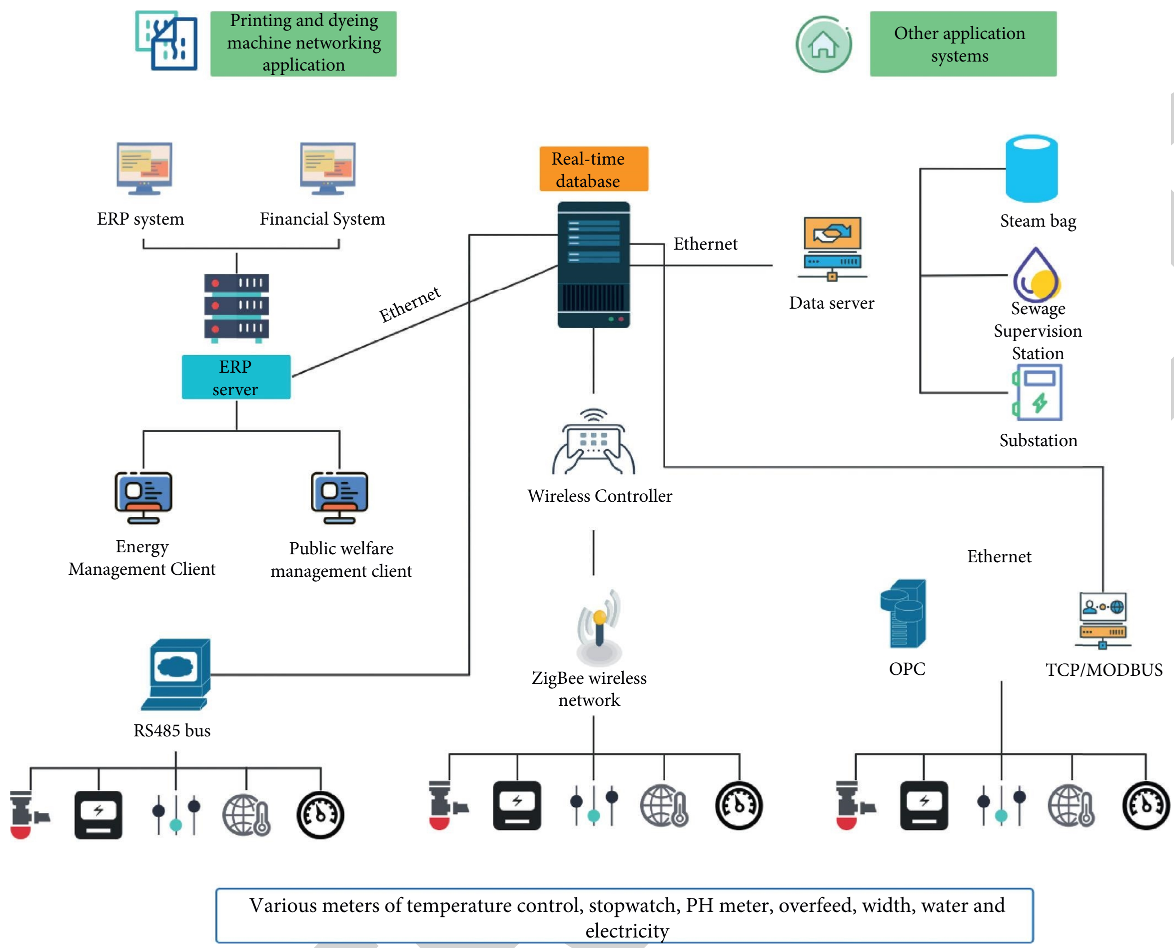


FIGURE 1: Machine learning algorithms.

enterprises and the progress of the industry require the improvement of machine learning algorithms to easily process huge amounts of information and data to achieve social development; thereby, the huge information data can be conveniently processed to realize the development of society [19]. Pasha and others proposed a machine learning algorithm through repeated experiments, which has great advantages in solving the current data mining problems. With the help of modern artificial intelligence technology and the modeling of various mathematical models, we can better use and solve the data problems [20]. Due to the relatively harsh production environment, the relevant sensor equipment is easy to damage under high temperature and high pressure conditions, resulting in some hesitation in intelligent exploration. There are also some obstacles in the upgrading of equipment. Considering the cost problem, most enterprises will not take too much risk to replace the equipment. Therefore, it is of great value to explore the feasibility of energy conservation and emission reduction from the perspective of data.

3. Methods

3.1. Data Preprocessing of the Setting Machine. This section mainly explains the improved neural network optimization method based on the gradient boosting tree model. According to the energy consumption model of the setting machine, the parameters of the model are adjusted to improve the generalization ability of the model. The specific steps are as follows.

Step 1. First, you need to set the initial value of the step size and the number of iterations. In general, initially choose a smaller step size for grid search to get the best number of iterations. Therefore, this paper just started by setting the step size to 0.1. Then, use the step size to perform grid search on the number of iterations and finally get a more suitable number of iterations.

Step 2. Then, it is necessary to determine the maximum depth of the decision tree and the minimum number of samples required for re-division of internal nodes and obtain the optimal solution of the two parameters through grid search.

Step 3. Then, it is necessary to adjust the minimum number of samples required for the re-division of the internal nodes and the minimum number of samples of the leaf nodes at the same time and obtain the optimal solution of the combination of the two parameters through grid search.

Step 4. Perform a grid search on the maximum number of features for feature sampling.

Step 5. Determine the proportion of subsampling by grid search and then perform sample sampling.

Step 6. Finally, the fitting ability and generalization ability of the energy consumption model of the training machine are further enhanced by comprehensively using the method of halving the step size and doubling the maximum number of iterations.

According to the needs of different processes, the equipment involved will also have certain differences; next, we will briefly introduce the printing and dyeing process; some devices with higher energy consumption are used. The first is the singeing machine used in the singeing process; usually, after the grey fabric enters the printing and dyeing factory, it first needs to go through the processes of inspection, turning over, batching, distributing blanks, and sewing heads; the matching blank is to join the front and back sides of the two grey fabrics on the same side; then, the singeing process is carried out, and the main energy consumed in the singeing process is natural gas. The second is the dye vat used in the printing and dyeing process; the main energy consumed in the dyeing process is water. Finally, there is the setting machine equipment used in the printing and dyeing setting process. The main purpose of setting is to eliminate the stress and strain accumulated in the fabric. After the setting process, the surface of the fabric becomes smooth and wrinkle-free, and the size is stable, and it has good thermal stability. For polyester fiber, improve its anti-wrinkle and non-ironing properties. The shaping process is mainly divided into three steps. (1) Grey fabric setting: it is also called presetting used to remove unfavorable impurities and fixation. The yellowing produced during this process can be removed during bleaching. Have to be aware of is that this step is not good for dyeability and high-quality requirements for grey fabrics. (2) Semifinished product shaping: it is also called medium shaping; it is easy to wrinkle in the process before shaping. The purpose of this step is to make the cloth surface smooth and reduce stains. Have to be aware of is that in the process, the dye adsorption is reduced, before and after mercerizing and high requirements for pretreatment. (3) Finished product finalization: Immediate finalization, this process will affect the dye sublimation fastness and make the dye change color. It must be noted that in this step, the yellowing caused by modeling cannot be removed, which requires a lot of dyes. In the heat setting process, the main energy consumed is electricity, natural gas, and a small part of water. We perform data preprocessing on each data table, including methods such as outlier and missing value processing, feature coding, data merging, feature screening, and removal of unique attributes; finally, the preprocessed feature data set is obtained [21, 22].

(1) Dealing with outliers and missing values: Outliers are also called outlier data; they can be understood as an unreasonable value in the data set. Judge whether a value is an outlier, it needs to be combined with the actual situation, such as the value of "−2"; if it represents the rotation speed of the circulating fan of the setting machine, this is obviously an abnormal value, but if it represents the local temperature (degrees Celsius) on a certain day, it is a reasonable value. In general, the processing methods for outliers include the following four: delete records containing outliers; mean value correction method to correct outliers; treat outliers as missing values, and it is processed by the missing value processing method; and it is not processed. The way we deal with data outliers when the setting machine is running is as follows: if the abnormal value related to the setting machine rarely occurs, select the method of deleting the abnormal value for processing. If there are many abnormal values of the setting machine, it indicates that there may be a problem with the sensor installed on the sizing machine and need to replace the relevant sensors to collect data again. For intermediate cases, the average value correction method is generally selected to deal with abnormal values [23–25]. The basic principle of the average value correction method will be briefly introduced below.

Assuming that a certain period of time is T, a certain runtime data value of the setting machine equipment at time t within T is x_t, if at some t_i, $(i = 1, 2, \ldots, n)$, moment in between the value of a certain process parameter of the setting machine equipment is abnormal and can use the calculation result of the following formula to replace the abnormal monitoring value:

$$x_{t_i} = \frac{\left(\sum_{t=t_1}^{t_{i-1}} x_t + \sum_{t=t_{i+1}}^{n} x_t\right)}{(n-1)}. \tag{1}$$

In the formula, $(n-1)$ is the period of time T, the number of a certain technological parameter value of the setting machine equipment.

Missing value refers to the situation where the value corresponding to a certain attribute is empty. In the case of very few and many missing values, in the same way as outliers, for intermediate cases, choose multiple exponential smoothing method to deal with and use three exponential smoothing methods to deal with missing values; the basic principle is as follows: for the missing value sequence of a certain process parameter of the setting machine equipment in a certain time period, according to the value of the process parameter of the setting machine equipment in the previous period of the missing value in this time period, on the basis of clarifying the length of the missing parameter sequence, the number of smoothing steps and data points is inserted; after clarifying the length of the missing parameter sequence, insert the number of smoothing steps and the number of data points [26, 27]. Use the following formula for smoothing:

$$\begin{cases} s_t' = ax_t + (1-a)s_{t-1}', \\ s_t'' = as_t' + (1-a)s_{t-1}'', \\ s_t''' = as_t'' + (1-a)s_{t-1}'''. \end{cases} \tag{2}$$

Among them, s_t', s_t'', and s_t'' are the values after primary, secondary, and smoothing, respectively; a is the weight coefficient of smoothing processing. In general, set $a = 0.5$. So, the smoothed value of the missing value is as follows:

$$x_{t+m}' = a_t + b_t m + \frac{1}{2} c_t m^2. \tag{3}$$

Among them, m represents the number of smoothing steps, set $m = 3$ here, a_t, b_t, and c_t are the coefficients of smoothing value, and the calculation formula is as follows:

$$\begin{cases} a_t = 3S_t' - 3S_t'' + S_t'', \\ b_t = \dfrac{a}{2(1-a)^2}\left[(6-5a)S_t' - (10-8a)S_t'' + (4-3a)S_t''\right], \\ c_t = \dfrac{a^2}{(1-a)^2}\left(S_t' - 2S_t'' + S_t''\right). \end{cases} \tag{4}$$

(2) Characteristic code: the printing and dyeing order data sheet contains the characteristic fabric name; there are 10 types of fabrics in total; they are four-sided stretch, crepe satin, fragrant cloud yarn, silk satin, chiffon beads, stretch satin jacquard, burnt-out velvet, jinlun yarn, tree pattern forging, and composite silk plain weave. For category data, one-hot encoding will be used for processing. One-hot encoding has a good effect in dealing with multi-classification problems. The following will briefly introduce the processing method of one-hot encoding. If the fabric name data are marked according to the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, although there is no problem with accuracy, there will always be some misunderstandings. For example, 2 is twice as large as 1, or if it has no special meaning, 9 is greater than 0. In order to avoid such misunderstandings, they can be expressed in a more equal matrix form.

(3) Data merging is based on the process parameter data sheet of the setting machine; merge data on it; according to the rotational speeds of the 12 circulating fans, the average rotational speed of the circulating fans is obtained. In the same way, find the average speed of the 3 exhaust fans and the average temperature of the 11-section oven; the process time is obtained by subtracting the process end time from the process start time. At the same time, according to the order number, the basic information of the order and the process parameters of the setting machine and the energy consumption of the setting machine is integrated. The initial feature data set is obtained so far [28].

(4) Feature screening determines the target column in the feature data set or constructs the target column by certain rules and analyzes the correlation between each data item in the initial feature data set and the target column. By analyzing the correlation coefficient, the data items with a correlation coefficient greater than 0.5% are screened out as the final feature data set.

(5) Remove the unique attributes, usually some id attributes; these attributes can only be used in the process of data merging. In the description of the distribution law of the sample itself, it is of little significance, so it should be simply deleted at the end. As you can see from the basic order information table, the attribute order number is the only attribute, so it will be deleted after the data are sorted and merged.

Difficulties in the Preprocessing of Production Data. The production data of different printing and dyeing enterprises are quite different in terms of the data format and the specific data content, which also brings the difficulty to the data preprocessing to a certain extent.

(1) *Sensor Data Acquisition Process.* Due to the harsh environment such as high temperature and high pressure in the process of printing and dyeing, the requirements for all kinds of sensors are relatively high. In the actual use process, the sensor is easy to damage. In the process of sensor update, there are also certain differences in the data acquisition accuracy.

(2) *Data Format and Standards.* Different printing and dyeing enterprises have a large distance in the process of intelligent transformation, so there are also great differences in the data format and standards. The first is that for a certain device, the relevant data should be collected. Secondly, there is the specific significance of the feature term represented for different feature data.

3.2. Comprehensive Energy Consumption Model per Unit Output of the Setting Machine. The relevant data of the shaping machine used in this paper come from a printing and dyeing factory, and the data come from the No. 1 shaping machine equipment of the enterprise. The energy consumption during the finalization process for a large extent is related to the volume of orders; for example, if you need to shape the fabric with an order of 30,000 meters, the energy consumed must be far greater than the same fabric with an order volume of 3 kilometers. Therefore, an indicator that can measure the energy consumption of different orders is needed. By constructing a comprehensive energy consumption model per unit output of the setting machine, measure the energy consumption of different orders. The comprehensive energy consumption model per unit output of the setting machine is established, and the comprehensive energy consumption model per unit output of the setting machine is used to generate the predicted value of the energy consumption category of the setting machine. Combined with the energy consumption category, the preprocessed characteristic data set is selected to obtain the training characteristic data set. The process of constructing the comprehensive energy consumption model per unit output of the setting machine is as follows: Extract energy-related data from the printing and dyeing sample data such as electricity consumption E, gas consumption G, water consumption W, and order volume data meters M. Through the formula of energy consumption per unit output of equipment, in order to calculate the energy consumption per unit output of a product, the formula is as follows:

$$P = \frac{P_q}{\sum N_{gh}}. \tag{5}$$

Among them, P_q represents the comprehensive energy consumption of the equipment, in kilograms of standard coal, and the calculation formula is as follows:

$$P_q = \sum_{i=1}^{n} \left(P_i * E_i\right), \tag{6}$$

where E_i represents the production activity, the consumption of type i energy in kind; P_i represents the conversion factor of standard coal for type i energy; and P represents the comprehensive energy consumption per unit output of the equipment, in kilograms of standard coal per 100 meters. Through calculation, the comprehensive energy consumption per unit output of the setting machine uses these data to represent the energy consumption of the setting machine, after entering the relevant data; the formula is as follows:

$$p = \frac{E * P_e + G * P_g + W * Pw}{M/100}. \tag{7}$$

For features whose values are continuous variables, calculate the Pearson correlation coefficient between the feature column and the target column and keep the feature columns whose correlation coefficient value is greater than 0.5%. For some ordinal variables or even-interval feature data that do not meet the assumption of normal distribution, calculate the Spearman correlation coefficient between the feature column and the target column and keep the feature columns whose correlation coefficient value is greater than 0.5%. The retained features are listed as the feature column for model training [29, 30]. The Pearson correlation coefficient calculation formula between two variables is as follows:

$$r_{xy} = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\left(\sqrt{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}\right)\left(\sqrt{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}\right)}. \tag{8}$$

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between rank variables; the original data are based on its average descending position in the overall data and were assigned a corresponding level. Eliminate the data consumption and the average circulating fan speed by processing the data. The 10 features used to predict the energy consumption classification of printing and dyeing shaping machines include cloth name, gas meter temperature, ambient temperature, gas meter pressure of No. 1 shaping, humidity at the front end, tail humidity, process time-consuming, vehicle speed, average speed of exhaust fan, and average temperature of drying room. The whole process of model establishment and implementation is shown in Figure 2.

4. Results and Analysis

4.1. Experimental Environment. Hardware platform is as follows: 16 GB, Quad-Core Intel Core i5@2.3 GHz.

Software platform is as follows: macOS Catalina 10.15.2, PyCharm 2019.2.3, Jupyter Notebook.

In this section, two types of group experiments are done. The first type is K-means unsupervised experiments, as well as supervised gradient boosting tree experiments, random forests, and approximate entropy classification experiments. The second type is regression experiment, including gradient boosting tree model, general decision tree model, and limit tree model. The experimental environment is as follows: hardware platform is 16 GB, Quad-Core Intel Core i5@2.3 GHz and software platform is mac OS Catalina 10.15.2, PyCharm 2019.2.3, Jupyter Notebook. For classification models, commonly used evaluation indicators include accuracy, precision, recall, and F1-score. In order to calculate these index data, you first need to understand several related values, as shown in Table 1.

With the four values in Table 1, we can calculate the accuracy, precision, recall, and F1-score values. The calculation formula and description are shown in Table 2.

For regression problems, Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Relative Error (MRE) are the main three algorithm evaluation indicators. Generally speaking, the smaller the MSE explains that the forecast data are better than the actual data, and the overall deviation is smaller. Same, the smaller the MAE and MRE, the better the prediction model. The RMSE is the square root of the sum of the square of the m ratio to the number of observations. It is used to measure the deviation between the observed value and the true value; the average absolute error is the average of the absolute error; it can better reflect the actual situation of the prediction value error; the standard deviation is the arithmetic square root which is used to measure the dispersion of a set of numbers itself. In Table 3, the calculation formula and description of each evaluation index are shown; the y that appears in the formula represents the actual data, y'represents predicted data, and L is the number of predicted data.

Through the gradient boosting tree model, regression analysis was performed on the data sample, the process parameters, and fabric types in the production process of the setting machine; data such as process time-consuming are used as input value; the calculated comprehensive energy consumption value is used as the output value, for training and performing prediction experiments on 1000 samples; on this basis, it is compared with the general decision tree model and the limit tree model. The contrast results of the regression model are shown in Figure 3.

As can be seen from Figure 3, in terms of predicting trends. The gradient boosting tree model is better than the decision tree model and the limit tree model in predicting the trend. In terms of predicting specific values, the gradient

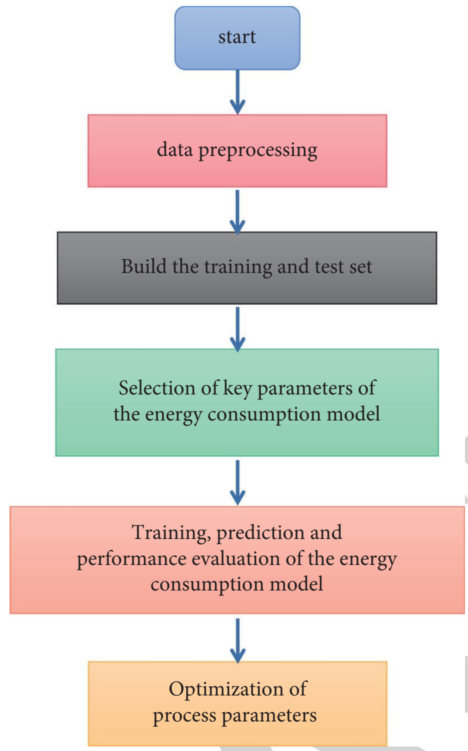


Figure 2: Model building and implementation process.

Table 1: Related values of algorithm evaluation index.

	Related, positive	Irrelevant, negative
Detected	True Positives TP, the positive class is judged to be a positive class	False Positives FP, negative category is judged as positive category, "preserve false"
Not detected	False Negatives FN, the positive class is judged as the negative class, "go to true"	True Negatives TN, negative category is judged as negative category

Table 2: Evaluation index of the classification model.

Algorithm evaluation index	Calculation formula	Description
Accuracy	Accuracy = (TP + TN/TP + FP + TN + FN)	Measure the proportion of accurate classification of all samples
Accuracy	Precision = (TP/TP + FP)	Also called precision rate, it measures the classification accuracy of positive samples, that is to say, the number of samples that are predicted to be positive samples, how many are really positive samples
Recall rate	Recall = (TP/TP + FN)	Indicates the proportion of positive samples correctly classified to the total positive samples
F1-score	F_1 = (2TP/2TP + FP + TN)	Harmonic average of precision rate and recall rate

boosting tree model is better overall. In the later stage of the energy consumption forecasting process of the enterprise, compared with the previous model, compared with the actual energy consumption, the gradient boosting tree model is better overall. By setting the gradient boosting tree model as a fitness function, the initial value of the learning rate is 0.1, and the number of iterations is 1000, and the optimization curves for different fabric types are obtained as shown in Figure 4.

Substituting the bestT, bestS, bestW, bestC, and bestP obtained by optimization into the model, compare its actual energy consumption per unit output with the optimized energy consumption per unit output, as shown in Figure 5; it is a comparison of energy consumption before and after optimization of Xiangyun yarn.

It can be found from the comparison before and after the energy consumption optimization of Xiangyun yarn that energy consumption has been reduced to a certain extent; in

Table 3: Evaluation index of the regression model.

Algorithm evaluation index	Calculation formula	Description
Mean square error	$\mathrm{MSE} = 1/L\sum_{i=1}^{L}\left(y_i - y_i'\right)^2$	The mean of the sum of squares of the difference between actual data and predicted data; this value is used to measure the “average error” of the forecast data.
Mean absolute error	$\mathrm{MRE} = \left(1/L\sum_{i=1}^{L}\left(y_i - y_i'\right)/y_i\right)$	The mean value of the sum of the absolute values of the difference between actual data and predicted data; the main advantage of this evaluation index is that it can overcome absolute and relative errors, the situation where positive and negative cancel each other out.
Average relative error	$\mathrm{MAE} = 1/L\sum_{i=1}^{L}\left\|y_i - y_i'\right\|$	The difference between the actual data and the predicted data, and the mean value of the ratio of the absolute value of the actual data, the evaluation index is mainly used to reflect the mean value of the percentage of the absolute error to the true value.

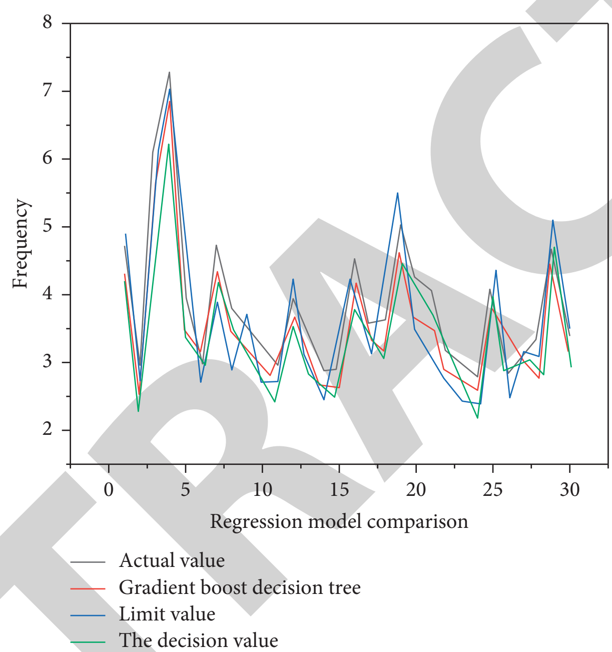


Figure 3: Comparison results of regression models.

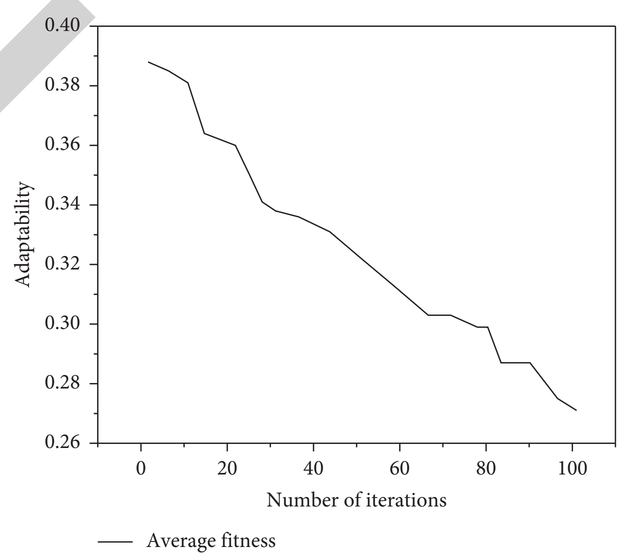


Figure 4: Optimization curve.

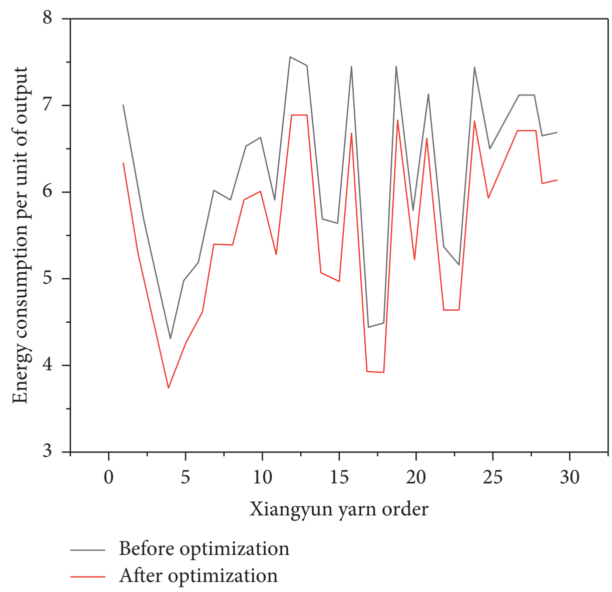


Figure 5: Comparison before and after energy consumption optimization of Xiangyun yarn.

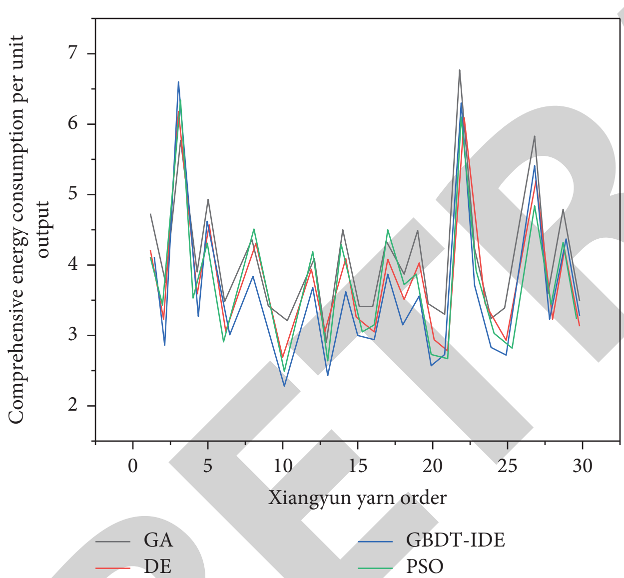


Figure 6: Comparison of optimization results of Xiangyun yarn.

order to a certain extent, the effect of energy saving and emission reduction has been achieved. In terms of optimization models, traditional genetic algorithms, differential evolution algorithms, and particle swarm optimization are used, compared with the proposed improvement based on the gradient boosting tree model and evolutionary algorithm for comparison. The comparison result is shown in Figure 6, which is the optimized result of Xiangyun yarn.

5. Conclusion

In recent years, people's pursuit of quality of life has become higher and higher, especially in terms of ecological environment. In addition, the country vigorously strengthens the construction of ecological civilization, during the inspection of the general secretary in Anji, Zhejiang and puts forward the idea that green water and green mountains are golden mountains and silver mountains. As a printing and dyeing textile industry with high energy consumption and high pollution, energy saving and emission reduction have become urgent problem to be solved in the industry. Initially, the optimization of the technological process and the optimization of the scheduling of the workshop achieved a certain effect to a certain extent. However, with the rise and maturity of Internet of things and big data technology, it is possible for us to explore energy-saving and emission reduction methods of printing and dyeing enterprises from the perspective of data. The printing and dyeing processes contain a large amount of production data, including process parameters and energy consumption data, and these large amounts of data also contain some unknown laws. Obtained through experimental results, the average absolute error of the limit tree is 0.01391, and the average relative error is 0.072461, the average absolute error of the decision tree is 0.01979, and the average relative error is 0.062579. Based on the data of printing and dyeing process, combined with order related information, process parameter related information, and energy-related multimedia information, the energy consumption category and energy consumption are constructed as the predicted value. After certain data preprocessing, the energy consumption of the subsequent production process is predicted by using the optimization algorithm for model training, and the process parameters are adjusted to a certain extent by using the prediction data, so as to achieve the effect of energy conservation and emission reduction. Energy consumption classification and regression method of printing and dyeing shaping process is studied. Starting from the data of the finalization process of the printing and dyeing process, combined with order-related information, process parameter-related information, energy consumption-related multimedia information, etc., the gradient lifting tree model is used for the classification and regression prediction of the shaping machine data, and the effect is compared with some other methods through comparative experiments to prove the effectiveness of the method.

This paper mainly explores the energy consumption optimization of printing and dyeing enterprises, and there are some reasons posed by industry restrictions. First of all, it is the current situation of most printing and dyeing industry; the table, line, and other equipment are easy to damage, data are inaccurate, and refinement is not in place. Some enterprises, especially small enterprises, still use the equipment in the state of ten or twenty years ago. Considering the cost problem, they do not introduce too many new equipment. Many orders are still made with old equipment. It leads to high energy consumption and more serious pollution. As a printing and dyeing enterprise with high energy consumption and high pollution, only by sharing some relevant data can scientific research really play a role. Only by truly paying attention to relevant scientific research can we explore the value as soon as possible to better improve the output and quality of the daily production of enterprises.

This paper mainly explores the data of energy consumption optimization of printing and dyeing enterprises, in which there are some problems caused by industry restrictions. The first is the current situation of most printing and dyeing industries. The tables, lines, and other equipment on the machine are easy to be damaged, the data are inaccurate, and the refinement is not in place. Some enterprises, especially small enterprises, still use the equipment in the state of more than ten or twenty years ago. Considering the cost, they do not introduce too many new equipment. A large number of orders are still produced with old equipment, resulting in high energy consumption and serious pollution. The second is the problem of data. As it involves enterprise production, the confidentiality of data is also very important. In terms of data security and confidentiality, most small enterprises cannot make a balance and can only stay in the more original production mode. The enterprises that have just begun to explore production data have problems such as inconsistent data format, which leads to certain limitations in exploring secondary energy conservation and emission reduction from data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] Z. Zhou, R. Xu, D. Wu, Z. Zhu, and H. Wang, "Illumination correction of dyeing products based on grey-edge and kernel extreme learning machine," *Optik*, vol. 127, no. 19, pp. 7978–7985, 2016.

[2] C. Satish, "jaya spider monkey optimization-driven deep convolutional lstm for the prediction of COVID-19," *Bio-Algorithms and Med-Systems*, vol. 16, no. 4, pp. 145–151, 2020.

[3] L. H. Zhang, "On optimizing the sum of the Rayleigh quotient and the generalized Rayleigh quotient on the unit sphere," *Computational Optimization and Applications*, vol. 54, no. 1, pp. 111–139, 2013.

[4] C. Liu, H. Lin, S. Gong, Y. Ji, and Q. Liu, "Learning topic of dynamic scene using belief propagation and weighted visual words approach," *Soft Computing*, vol. 19, no. 1, pp. 71–84, 2015.

[5] S. Narula, K. Shameer, A. M. Salem Omar, J. T. Dudley, and P. P. Sengupta, "Machine-learning algorithms to automate morphological and functional assessments in 2d echocardiography," *Journal of the American College of Cardiology*, vol. 68, no. 21, pp. 2287–2295, 2016.

[6] P. Charonyktakis, M. Plakia, I. Tsamardinos, and M. Papadopouli, "On user-centric modular qoe prediction for voip based on machine-learning algorithms," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1443–1456, 2016.

[7] P. Plawiak, T. Sosnicki, M. Niedzwiecki, Z. Tabor, and K. Rzecki, "Hand body language gesture recognition based on signals from specialized glove and machine learning algorithms," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1104–1113, 2016.

[8] H. Y. Tsao, P. Y. Chan, and E. C. Y. Su, "Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms," *BMC Bioinformatics*, vol. 19, no. S9, pp. 283–121, 2018.

[9] J. A. Atkinson, G. Lobet, M. Noll, P. E. Meyer, M. Griffiths, and D. M. Wells, "Combining semi-automated image analysis techniques with machine learning algorithms to accelerate large scale genetic studies," *GigaScience*, vol. 6, pp. 1–7, 2017.

[10] V. Prasad, T. S. Rao, and M. S. P. Babu, "Thyroid disease diagnosis via hybrid architecture composing rough data sets theory and machine learning algorithms," *Soft Computing*, vol. 20, no. 3, pp. 1179–1189, 2016.

[11] F. Azmat, Y. Chen, and N. Stocks, "Analysis of spectrum occupancy using machine learning algorithms," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 9, pp. 6853–6860, 2016.

[12] G. Bartsch, A. P. Mitra, S. A. Mitra et al., "Use of artificial intelligence and machine learning algorithms with gene expression profiling to predict recurrent nonmuscle invasive urothelial carcinoma of the bladder," *The Journal of Urology*, vol. 195, no. 2, pp. 493–498, 2016.

[13] C. Stoica, J. Camejo, A. Banciu et al., "Water quality of danube delta systems: ecological status and prediction using machine-learning algorithms," *Water Science and Technology*, vol. 73, no. 10, pp. 2413–2421, 2016.

[14] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, and K. L. Zimmerman, "Evaluation of supervised machine-learning algorithms to distinguish between inflammatory bowel disease and alimentary lymphoma in cats," *Journal of Veterinary Diagnostic Investigation*, vol. 28, no. 6, pp. 679–687, 2016.

[15] J. Liu, Y. Ye, C. Shen, Y. Wang, and R. Erdelyi, "A new tool for cme arrival time prediction using machine learning algorithms: cat-puma," *The Astrophysical Journal*, vol. 855, no. 2, p. 109, 2018.

[16] M. A. H. Khan, B. Thomson, R. Debnath, A. Motayed, and M. V. Rao, "Nanowire-based sensor array for detection of cross-sensitive gases using pca and machine learning algorithms," *IEEE Sensors Journal*, vol. 20, no. 11, pp. 6020–6028, 2020.

[17] A. R. Verma, B. Gupta, and C. Bhandari, "A comparative study of ecg beats variability classification based on different machine learning algorithms," *Augmented Human Research*, vol. 5, no. 1, pp. 16–18, 2020.

[18] I. Mandal and S. Pal, "Modelling human health vulnerability using different machine learning algorithms in stone quarrying and crushing areas of Dwarka river Basin, Eastern India," *Advances in Space Research*, vol. 66, no. 6, pp. 1351–1371, 2020.

[19] M. Roimi, A. Neuberger, A. Shrot, M. Paul, and Y. Geffen, "Early diagnosis of bloodstream infections in the intensive care unit using machine-learning algorithms," *Intensive Care Medicine*, vol. 46, no. 3, pp. 454–462, 2020.

[20] A. R. Olivera, V. Roesler, C. Iochpe et al., "Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes - elsa-brasil: accuracy study," *Sao Paulo Medical Journal*, vol. 135, no. 3, pp. 234–246, 2017.

[21] R. Rama Devi and M. Abualkibash, "Intrusion detection system classification using different machine learning algorithms on kdd-99 and nsl-kdd datasets—a review paper," *International Journal of Computer Science and Information Technology*, vol. 11, no. 03, pp. 65–80, 2019.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9869657, 1 page
https://doi.org/10.1155/2023/9869657

Retraction

Retracted: A Novel Literary Translation Text Classification Method Based on Distributed Incremental Sequence Data Mining Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Sun, "A Novel Literary Translation Text Classification Method Based on Distributed Incremental Sequence Data Mining Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 1656151, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1656151, 10 pages
https://doi.org/10.1155/2022/1656151

Research Article

A Novel Literary Translation Text Classification Method Based on Distributed Incremental Sequence Data Mining Algorithm

Ji Sun

Zhejiang International Studies University, Hangzhou 310023, China

Correspondence should be addressed to Ji Sun; sunji@zisu.edu.cn

Received 13 March 2022; Revised 6 April 2022; Accepted 15 April 2022; Published 9 May 2022

Academic Editor: Chin-Ling Chen

With the rapid development of electronic information technology and Internet technology, people's ability to generate and collect data is also increasing. With the rapid development of international exchanges, the number of literary translation texts has also increased dramatically. The information contained in the huge data of literary translation texts is huge, but these data are disorganized at this stage. Therefore, the classification of literary translation texts has become the key to efficient management of translated text information. Literary translation text classification is the process of classifying a given text as one or several of several predetermined text categories according to the content of the text. As one of the key steps in processing huge amounts of text data, text classification is generally regarded as the orderly organization of text sets, that is, grouping similar and related texts together. In this way, the problem of information clutter can be solved to a greater extent, and the efficiency of users' discovery, filtering, and analysis of text information resources can be effectively improved. At present, the basis of text classification is mainly based on the characteristics of words in the article, to analyze the correlation between words and categories. This approach ignores information such as word order and collocations in literary translation texts. To solve this problem, this paper introduces the distributed incremental sequence data mining algorithm into the classification of literary translation texts. This method can fully mine the features of syntactic order and other features based on considering the words and phrasing characteristics of articles in different application fields. The text classification effect is strengthened by discovering more effective information. The experimental results show that the method can improve the classification performance of literary translation texts.

1. Introduction

With the advancement of globalization, international exchanges and learning have become more common, and the dissemination of literary texts has grown in scope. As the world's primary language of communication today, English has gradually produced translated texts from various literature. Many literary translation texts were created over time. When confronted with a massive amount of literary translation text information, knowing how to efficiently filter useful information, accurately locate information categories, and deftly explore hidden information has become increasingly important. Based on this demand, text classification technology emerges as needed, having a significant impact on the process of knowledge discovery. The early text classification was done entirely by hand, and the work was arduous, clumsy, time consuming, labor intensive, and inefficient. Until the late 1950s, the IBM Corporation of the United States led the way in incorporating the concept of word frequency statistics into the text classification process, with positive results. This also marks the beginning of a new phase in text classification research. Many experts and scholars have conducted extensive and in-depth research on text classification methods since then, and many text classification methods have been proposed.

Because of Internet's rapid development, all types of information are flooding the network, and a keyword search in a search engine will return millions of results. The ability to use effective classification methods to filter useful category information has a wide range of applications in real life [1–3]. Many researchers from various fields have conducted extensive research on this topic, and various text classification methods have emerged [4, 5]. Text classification based on similarity is a widely used method in the field of text classification [6–8]. Several methods for calculating vector similarity have been proposed, including Euclidean distance, Jaccard similarity, cosine similarity, Kullback–Leibler divergence, Canberra distance metric, hammering distance, dice coefficient, and pairwise adaptive, among others. Measures of similarity have been developed. It is commonly used in text classification [9, 10]. This different relationship, however, is a significant factor influencing classification performance. The use of bag-of-words is also very mature in the field of text classification [11–13]. A document is frequently represented as a vector, with each component containing the value corresponding to the corresponding feature. This feature value could be word frequency, associated word frequency, or something else. Many studies have begun to conduct in-depth research on text lexical rules in order to improve the performance of text classification [14]. Because of their simplicity, word-based methods are preferred by researchers in a variety of fields. At this point, many word-based feature selection methods have been proposed by various scholars [15–17]. However, word-based methods usually ignore the association between words, which will lead to loss of information about paragraphs and sentences in the text and will easily cause problems such as polysemy, synonyms, and noise interference [18].

Texts in various application fields, in general, have their own distinct writing characteristics, or literary habits. Some of these habits are reflected in the article's wording and phrasing, while others are reflected in word collocation, syntactic order, or other aspects. Because of the presence of these characteristics, we can easily identify texts in a variety of application fields. As a result, whether it is the initial manual recognition or machine recognition, the information used for text classification is likely to belong to the article characteristics of the application field contained in each document. Most current text classification methods primarily complete the task of text classification by identifying, extracting, and utilizing textual characteristics in various fields. The current problem is that the traditional classification method still focuses on the relevance of the article and only determines which feature words correspond to the field. However, words are relatively isolated from each other, and the characteristics of word collocation and syntactic order are not considered, so the information that can be extracted is relatively limited. Sequence pattern mining can dig out the pattern sequences hidden in the transaction set, and all the elements in these pattern sequences are frequent and relatively sequential. This order can be relative to time or relative to space. This shows that the sequential pattern mining method can fully and effectively utilize the sequential characteristics of transactions. In view of the shortcomings of the existing text classification methods, this paper attempts to introduce the idea of sequential pattern mining into the text classification method and uses the distributed incremental sequential data mining algorithm to propose a method of literary translation text based on the distributed incremental sequential data mining algorithm. The new method can fully mine text syntactic sequence features such as word order and word collocation based on considering the characteristics of words and phrasing of articles in different application fields. Thereby, more effective information can be discovered to strengthen and improve the effect of text classification.

2. Knowledge Based on Sequence Patterns

2.1. Sequence Pattern Data Mining. The goal of sequential pattern mining [19] is to find all sequential patterns in a sequence database that have a frequency greater than a predefined threshold. Mining for sequence patterns is analogous to mining for association rules. However, the former is more concerned with the temporal or spatial dependencies of event elements than the latter, i.e., the former is more concerned with event sequence characteristics. This order can be both temporal and spatial in nature. The original goal of sequence mining was to uncover sequences of frequent buying patterns in a transaction database with transaction times. In this method, we can determine the purchase behavior trajectory of most clients over a certain time period. Later, with the deepening of algorithm research and the upgrading of technology, especially the optimization of mining algorithms and the use of high-performance computers, the application field of sequential pattern mining has become very broad. At present, it is widely used in professional fields such as customer purchasing behavior prediction, medical diagnosis, web page access pattern prediction, industrial control, and gene sequence analysis.

Sequential pattern mining is one of the important technologies in the field of data mining and has applications in many fields. At present, there are a lot of related research studies on sequential pattern mining algorithms, and the proposed algorithms are also very comprehensive. According to different mining strategies, sequential pattern mining algorithms can be divided into two categories, one is the algorithm based on the breadth-first search strategy, and the other is the algorithm based on the depth-first search strategy. Among them, the representative algorithms based on the breadth-first search strategy are the Apriori-All algorithm [20], the GSP algorithm [21], and the SPADE algorithm [22]. Typical depth-first search algorithms mainly include FreeSpan algorithm [23], PrefixSpan algorithm [24], and SPAM algorithm [25].

2.2. Text Classification Based on Sequential Patterns. As illustrated in Figure 1, text classification based on sequential patterns consists mostly of two stages. The first task is to create a library of classification patterns. The sequential pattern mining approach is used to mine the sequential patterns belonging to this category for each batch of training texts from a known category, resulting in the formation of the pattern sublibrary of this category. After each category's sublibraries are generated, they are aggregated into the final pattern library $P = \{P_1, \ldots, P_j, \ldots, P_k\}$. The next step is to examine the freshly entered text. The input text sequence is

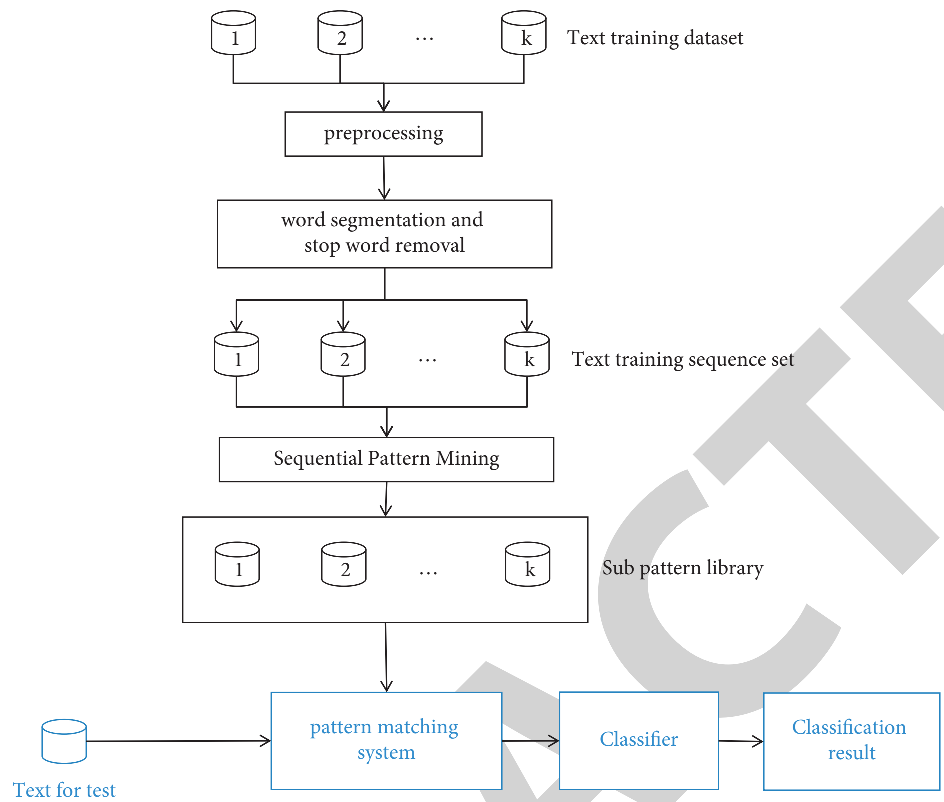


Figure 1: Architecture diagram of text classification based on sequence pattern.

generated by turning each text into a line of text characters using related approaches. Second, pattern matching is carried out. Finally, the text is classified in accordance with the classification principle. The classification of literary translation texts in this paper is based on the framework shown in Figure 1.

According to Figure 1, it can be concluded that the text classification process based on sequence patterns is as follows:

Text preparation is the first. Only discrete data can be analyzed with sequential pattern mining algorithms, and it is not always evident whether a dataset is discrete. If the data are continuous, it must first be discretized. To obtain a set of word sequences for each category, we first store all the texts in the training text set individually according to distinct categories and then execute word segmentation and stop word removal processing on them according to a unified standard.

Second, each category's frequent sequence patterns are mined to create a categorization pattern library. Each category's set of word sequences is turned into a standard set of transactions with timestamps and sequence IDs. The sequential pattern mining method is then utilized for the standard transaction sets of different categories to uncover the frequent patterns belonging to each category, and the subpattern library belonging to each category is obtained. The categorization pattern library is then obtained.

Finally, text classification is carried out. The new text is input to be categorized, and pattern matching is executed with all frequent patterns in the classification pattern library that has been created. We count the number of common patterns that match the new text in each subpattern library. Finally, documents having unidentified category attribute values are categorized using the definition classification principle.

3. Distributed Incremental Sequence Data Mining Algorithm

3.1. Distributed Computing Framework MapReduce. A sequential pattern mining method based on the distributed lexical sequence tree algorithm was proposed in reference [26]. Based on this research, this work employs a MapReduce-based distributed incremental sequential pattern mining approach. In a huge data setting, the approach can be utilized to overcome the incremental maintenance problem of sequential pattern mining. To deal with huge data challenges, MapReduce employs a distributed programming framework that employs a divide-and-conquer method. The system conceals the internal mechanics of data segmentation and distribution, task scheduling, intermachine communication, and fault tolerance from the programmer. It enables inexperienced programmers to efficiently manage the system's resources in a distributed system. The MapReduce approach not only simplifies distributed programming, but it also allows for more efficient processing of big data volumes.

Our method is divided into two MapReduce stages. The first stage consists primarily of reading the sequence's input split and determining the support count of the frequent 1 itemsets, as well as determining whether the frequent 1 itemsets belong to the incremental dataset using flag variables. The CMP data structure is created in the second stage. The candidate data are generated by using the CMP data structure and backward expansion. Furthermore, the prepruning attributes used in backward mining prevent erroneous candidate sequences from being generated in the input database, which speeds up the mining process.

3.2. Mining Frequent 1 Sequence. The first stage is used to mine frequent 1 sequence. Each Map1 identifies the input sequence dataset and determines whether item x belongs to the corresponding sequence's incremental dataset IDB. Item x stores 1 in distributed cache F if item belongs to IDB, otherwise 0. Then, Map outputs item x and its matching value in F as <key, value> and Reduce1<x, value> as input. The item count and flag variables are set to zero. The function of the flag variable is to indicate whether the item is included in the IDB. Reduce1 is used to determine how many values are related to each item. Reduce1 sets the flag variable to 1 after getting the item value 1, which aids in finding the incremental union in the second stage. At last, as the first stage's output, the frequent 1 itemsets, and their related counts and flags are used. Figure 2 depicts the first stage of this method's procedure.

3.3. Mining Frequent k Sequences. First, the frequent 1 itemsets, minimum support counts, and flags in the original dataset output from the first stage are used as the input of the second stage. Second, Map2 creates incremental unions and constructs CMP data structures, and uses prepruning attributes for backward mining. Finally, Reduce2 is used to mine frequent k sequences. The second stage process is shown in Figure 3.

3.3.1. Discovering the Incremental Union. The distributed cache file is read in the second stage, and the incremental union is found. For each item in the distributed cache file, we check if its flag is equal to 1. If it is equal to 1, we join the item to the incremental union, otherwise the 1 mode cannot be added to the incremental union.

3.3.2. Constructing the CMP Data Structure. This paper constructs the CMP data structure to optimize the speed of the algorithm. CMP(i) construction: By definition, CMP(i) is a mapping of an item and its preceding item list (co-occurrence list, CLST) relative to the extension of the itemset. The sequence scans from the last itemset to the first to match the CLST. If there is corresponding item to εi exists in CMP(i), then the CLST of εi is retrieved from CMP(i) of εi, and the co-occurrence item εj is checkedcby the retrieved CLST to find its count. If item j did not previously exist in CLST, it is included in CLST (εi). If the entry in CMP(i) corresponding to term i does not exist, then a CLST for term i is created by including the co-occurrence term εj. The CMP(i) of εi is updated with the corresponding CLST CMP(s) construction: By definition, a CMP(s) is an item mapping to its preceding list relative to the sequence expansion, and creating a CMP(s) is basically the same as creating a CMP(i).

3.3.3. Backward Mining Algorithm. The input of the backward mining algorithm is the sequence, the end projection (Epj) and H projection (Hpj) of the CMP(i), and CMP(s) sequences of the first item of the sequence, and the output is the updated data set frequent sequential patterns in UDB. The 1-mode end projection is the Epj from the backward extension. The sequence's backward extension is inferred by the CMP(i) and CMP(s) of the sequence. If the CMP(i) of the sequence does not exist, then itemset extension does not exist. The Epj of the generated sequences can be found in the Epj of the sequences, and the H projections of the sequences are scanned to find the H projections of the sequences extending from them. Support counts for these sequences are the sum of Epj's and Hpj's size. When the support counts of all extensions meet the lower threshold, they will be added into frequent sequences list and be called with recursion.

3.3.4. Generating Candidate Sequences. After creating the CMP data structure, a candidate set is generated. This paper designs two candidate generation rules based on backward mining to avoid generating misleading data so as to generate sequences of candidate efficiently. We assume the following definitions:

(1) For a given sequence pattern of length k $s = [\varepsilon_k \varepsilon_{k-1} \ldots \varepsilon_1]$, generate itemset candidates from a extensions that belong to CMP(i) (b) $C_{k+1}^{\varepsilon} = [\{a\varepsilon_k\}\varepsilon_{k-1} \ldots \varepsilon_1]$, where b is the first term in ε_k

(2) For a given sequence pattern of length k $s = [\varepsilon_k \varepsilon_{k-1} \ldots \varepsilon_1]$, generate sequence candidates by extending a that belong to CMP(s) (b) $C_{k+1}^{s} = [\{a\varepsilon_k\}\varepsilon_{k-1} \ldots \varepsilon_1]$, where b is the first term in ε_k

3.3.5. Generating Prepruning Sequences. Computing the support counts is quite slow. Mining speed can be improved if the generated candidates are pruned before computing support counts. This paper defines prepruning properties based on the CMP data structure. Sequences are trimmed early if their length exceeds a certain value, according to their properties. As shown in Figure 4, the following definitions are assumed:

3.3.6. Algorithm Implementation. The resources used are initialized through the SETUP function in the second phase of the MapReduce platform in this paper. After getting the input, Mappers treats infrequent items with pruning, then creates CMP(i) and CMP(s), and finds Projection and end projection of each 1-mode a. Each 1-mode is checked to see if it belongs to an incremental union to get the 1-modes of stable and expand them to seek the frequent sequences. If 1 mode is unstable, the function of backward mining will be

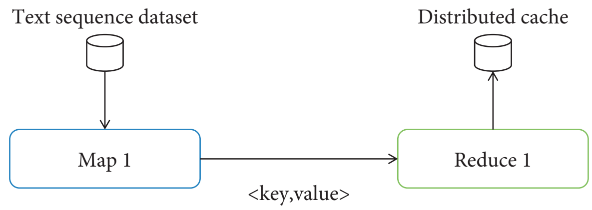


Figure 2: Flowchart of the first stage.

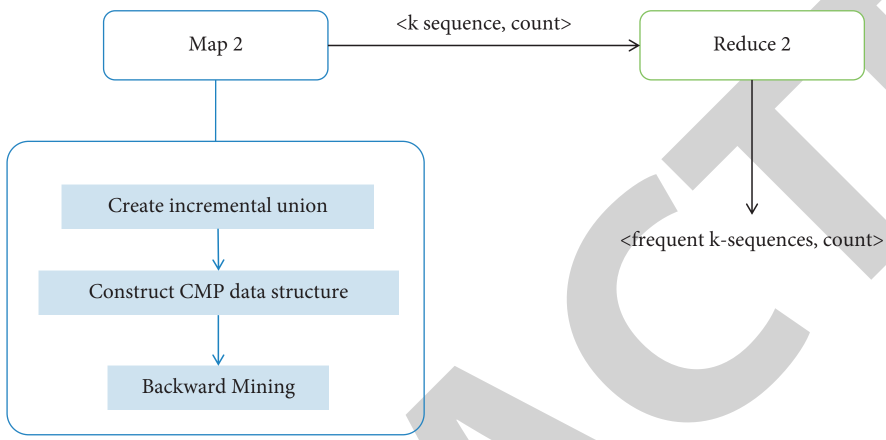


Figure 3: Flowchart of the second stage.

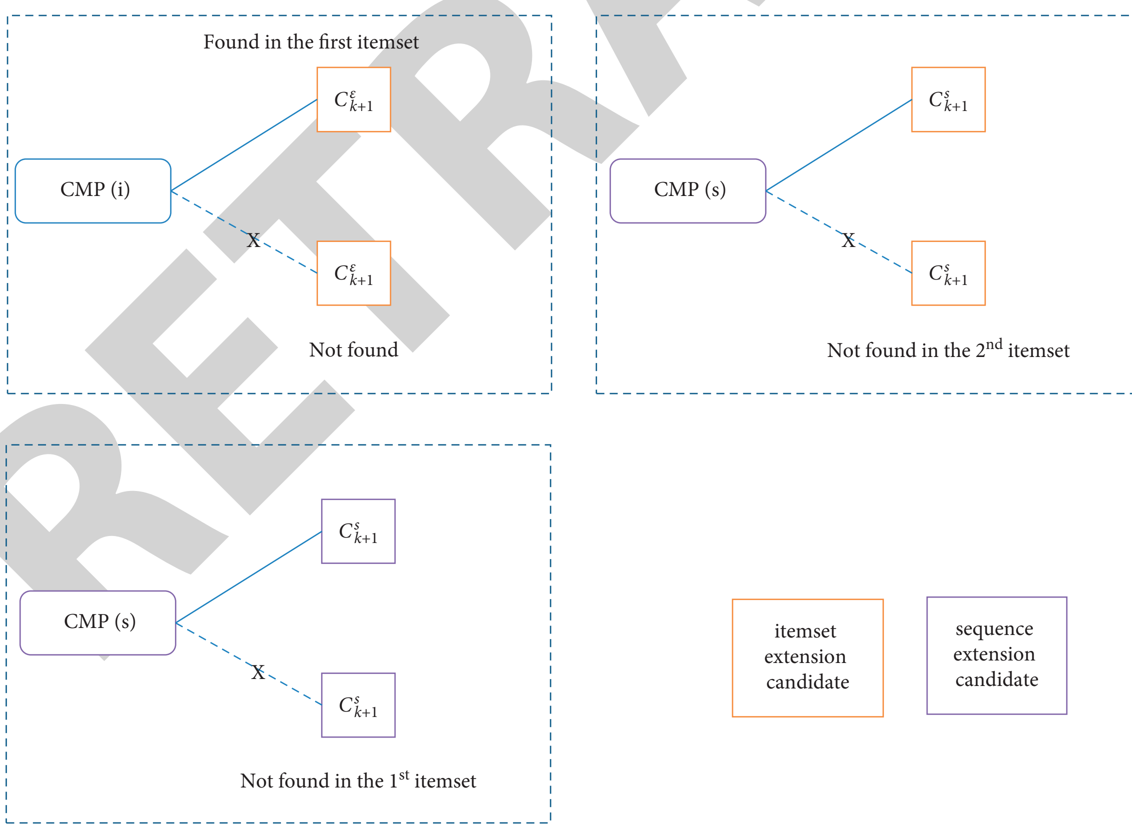


Figure 4: Definitions of prepruning.

called for each unstable ones, and 5 parameters are passed to the function when calling: unstable 1-mode a, the extended sequence of a, the itemset extended sequence of a, the end projection of a, and the H projection of a, where the H projection of a is calculated from the difference between the projection of a and the end projection of a.

Reduce2 takes k and value v as input and initializes the item count to 0. Reduce2 computes the items' values, after receiving the value about the item, and increments the item count. If the item count is greater than the product of the min support count and $|UD|$, we output k and item counts and then write them to files in the Hadoop distributed file system HDFS.

4. Experiment

4.1. Experimental Design. The hardware environment used in the experiment in this paper is CPU frequency 3.40 GHz; memory 4G. Software environment: Operating system: Windows10; IDE : MyEclipse9.0; Development environment: JDK 1.6; Development language: Python 2.7; Database: MySql.

The evaluation metrics used in the experiment are Recall (R), Precision (P) and F1. Recall refers to the ratio of the number of samples correctly judged by the classifier to the total number of samples belonging to this class; precision refers to the proportion of samples that truly belong to this class among the samples judged by the classifier as this class. We assume that when retrieving documents from a large-scale document collection, the documents can be divided into four groups: A represents relevant documents retrieved by the system; B represents irrelevant documents retrieved by the system; C represents relevant but not retrieved by the system documents; D stands for irrelevant documents that have not been retrieved by the system. Then, the text classification recall and precision can be defined as

$$\begin{aligned} \text{Recall} &= \frac{A}{A+C} \times 100\%, \\ \text{Precision} &= \frac{A}{A+B} \times 100\%, \\ F1 &= 2 \times \text{Recall} \times \frac{\text{Precision}}{(\text{Recall} + \text{Precision})}. \end{aligned} \tag{1}$$

In general, the text classification process can be divided into two parts: training the model and classifying the text. The main goal of training is to construct a classification model using the special relationship between text features and text categories through a set of training texts of known categories. The generalized model training process includes five steps: acquisition of training text sets, text preprocessing, text feature extraction, text representation, and selection of classification algorithms to construct classifiers. Categorizing text is to use the classifier obtained from the above training results to classify new and unknown text into categories, and the process of "sticking" category labels, mainly including text preprocessing, text representation, text classification, and classification performance evaluation. The main feature selection methods used in this step are mutual information (MI) and information gain (IG). In order to compare the performance of text algorithms, the comparison classification algorithms used include classic support vector machines (SVM) and random forests (RF).

4.2. Experimental Data. The datasets used in the experiments are two public corpora: Reuters-21578 and 20-Newsgroups.

4.2.1. Reuters-21578. The Reuters-21578 corpus is widely used for text mining. The data, originally collected by Carnegie Corporation and Reuters newsgroups, contained 21,578 documents on 135 different topics. The experiment uses a total of 9,980 documents from the top 10 topics. Because CORN and WHEAT are closely related to GRAIN, they are classified into GRAIN, so they are also called R8. This paper uses the 10-fold cross-validation method to divide the training dataset and the test dataset. We choose 80% as the training set and 20% as the test set. Table 1 describes the categories of documents in the dataset Reuters-21578.

The 20-Newsgroups dataset contains 4 major categories: COMP, REC, SCI, and TALK; each category contains 4 subcategories, with a total of 15,033 documents. Table 2 describes the document descriptions for each category in the dataset 20-Newsgroups. We choose 80% as the training set and 20% as the test set.

4.3. Experimental Results. The experimental results obtained by each classification algorithm on the Reuters-21578 dataset are shown inTable 3–5.

From the experimental data in Table 3–5 and the comparison chart of the classification results of each method shown in Figure 5, the following experimental conclusions can be drawn:

(1) Regardless of the classification method, the experimental results obtained by the IG-based feature extraction method are generally better than those obtained by the MI-based feature extraction method. This is because the IG-based feature extraction method can perform global feature extraction, and the extracted features are often valid for all classes.

(2) When the IG feature extraction method is used, the classification performance of the method proposed in this paper is better than other methods in most cases. The classification performance of RF is not much different from that of the method used in this paper, especially in the text classification of the first 6 classes. When using the MI feature classification method, comparing the experimental results obtained by different classifiers, in the first four categories of text classification, the method used in this paper is close to the classification performance of RF and the performance in the latter four categories. In this paper, the method used has distinct advantages. From the beginning to the end, the classification performance shown by the SVM method is not very good.

(3) Based on the above analysis, the method proposed in this paper can obtain better results than other classifiers no matter which feature extraction method is used. This fully shows that the method in this paper has certain advantages in the similar methods. In addition, according to the size of the specific data in each table, the text classification accuracy obtained by this method

Table 1: Dataset Reuters-21578 details.

Category	Total of each category	Training number	Testing number
ACQ	2369	1895	474
CRUD	578	462	116
EARN	3964	3171	793
GRAIN	582	466	116
INTEREST	478	382	96
MONEY	717	574	143
SHIP	286	229	57
TRADE	486	389	97

Table 2: Dataset 20-Newsgroups details.

Category	Total of each category	Training number	Testing number
COMP	3870	3096	774
REC	3968	3174	794
SCI	3945	3156	789
TALK	3250	2600	650

Table 3: Recall values on the Reuters-21578 dataset.

Method	Feature extraction	ACQ	CRUD	EARN	GRAIN	INTEREST	MONEY	SHIP	TRADE
SVM	IG	0.4835	0.4909	0.6536	0.5986	0.5069	0.5985	0.4943	0.4665
	MI	0.4328	0.4251	0.5902	0.5265	0.4401	0.5336	0.4438	0.4240
RF	IG	0.5002	0.8418	0.7420	0.8910	0.7084	0.6894	0.8165	0.5900
	MI	0.4301	0.7577	0.6436	0.7970	0.6262	0.5988	0.6815	0.5172
Proposed	IG	0.5976	0.8417	0.8574	0.8884	0.9608	0.824	0.9061	0.8217
	MI	0.5395	0.7306	0.6960	0.7303	0.8612	0.802	0.8105	0.6827

Table 4: Precision values on the Reuters-21578 dataset.

Method	Feature extraction	ACQ	CRUD	EARN	GRAIN	INTEREST	MONEY	SHIP	TRADE
SVM	IG	0.5313	0.5103	0.6611	0.6224	0.5574	0.6072	0.4982	0.5290
	MI	0.4568	0.4294	0.6148	0.5315	0.4473	0.5997	0.4972	0.4470
RF	IG	0.5109	0.8771	0.7785	0.9149	0.7149	0.7164	0.8648	0.6454
	MI	0.4471	0.7588	0.6517	0.8007	0.6884	0.6223	0.7236	0.5617
Proposed	IG	0.6692	0.9135	0.9370	0.9288	0.8932	0.8821	0.9645	0.8534
	MI	0.5452	0.7463	0.7226	0.7912	0.8539	0.810	0.8671	0.6876

Table 5: F1 values on the Reuters-21578 dataset.

Method	Feature extraction	ACQ	CRUD	EARN	GRAIN	INTEREST	MONEY	SHIP	TRADE
SVM	IG	0.5063	0.5004	0.6573	0.6103	0.5310	0.6028	0.4962	0.4958
	MI	0.4445	0.4272	0.6022	0.5290	0.4437	0.5647	0.4690	0.4352
RF	IG	0.5055	0.8591	0.7598	0.9028	0.7116	0.7026	0.8400	0.6165
	MI	0.4384	0.7582	0.6476	0.7988	0.6558	0.6103	0.7019	0.5385
Proposed	IG	0.6314	0.8761	0.8954	0.9082	0.8555	0.8512	0.9344	0.8373
	MI	0.5423	0.7384	0.7091	0.7595	0.8320	0.7680	0.8378	0.6851

generally exceeds 0.8, which shows that the text classification based on this method has certain practicability.

The experimental results obtained by each classification algorithm on the 20-Newsgroups dataset are shown in Table 6:

From the experimental data in Table 6 and the comparison chart of the classification results of each method shown in Figure 6, the following experimental conclusions can be drawn:

(1) The 20-Newsgroups dataset has 4 categories, and the experimental data of the first 3 categories shown in

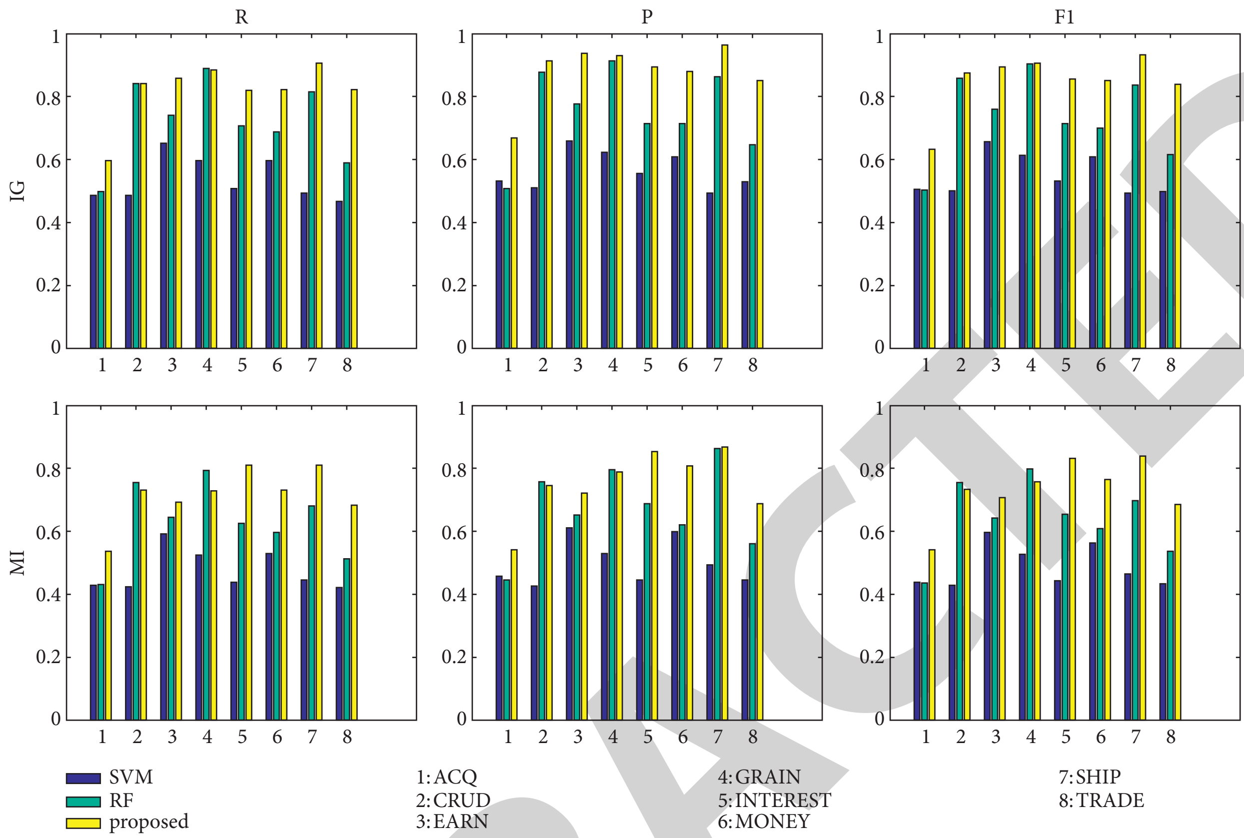


FIGURE 5: Comparison of experimental results on the Reuters-21578 dataset.

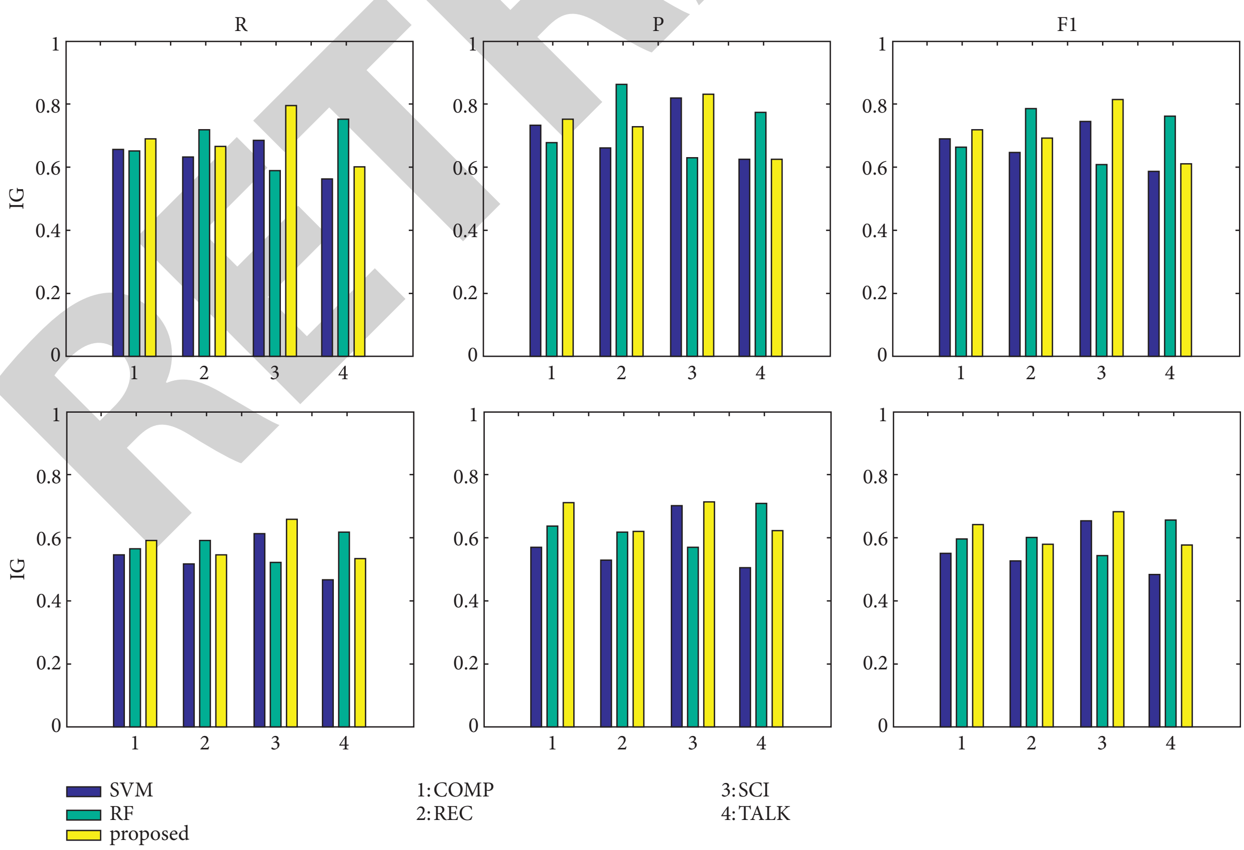


FIGURE 6: Comparison of experimental results on the 20-Newsgroups dataset.

Table 6: Experimental results on the 20-Newsgroups dataset.

Method	Category	Feature extraction	Recall	Precision	F1
SVM	COMP	IG	0.6555	0.7349	0.6929
		MI	0.5445	0.5700	0.5569
	REC	IG	0.6294	0.6624	0.6455
		MI	0.5153	0.5284	0.5217
	SCI	IG	0.6868	0.8192	0.7472
		MI	0.6093	0.7016	0.6522
	TALK	IG	0.5603	0.6284	0.5924
		MI	0.4654	0.4992	0.4817
RF	COMP	IG	0.6534	0.6816	0.6672
		MI	0.5609	0.6363	0.5962
	REC	IG	0.7180	0.8670	0.7855
		MI	0.5924	0.6186	0.6052
	SCI	IG	0.5904	0.6268	0.6081
		MI	0.5244	0.5713	0.5469
	TALK	IG	0.7536	0.7723	0.7628
		MI	0.6180	0.7087	0.6602
Proposed	COMP	IG	0.6923	0.7550	0.7223
		MI	0.5888	0.7105	0.6440
	REC	IG	0.6668	0.7271	0.6957
		MI	0.5471	0.6205	0.5815
	SCI	IG	0.7989	0.8315	0.8149
		MI	0.6561	0.7127	0.6832
	TALK	IG	0.6012	0.6266	0.6136
		MI	0.5341	0.6204	0.5740

Table 6 show that the method used in this paper has the best classification performance, regardless of whether it is based on IG or MI. The classification performance of RF is not stable. The performance of the SVM algorithm is stable, but the overall classification performance is lower than the method proposed in this paper.

(2) On the 20-Newsgroups dataset, the difference between the classification results obtained by the two feature extraction methods becomes smaller. However, the experimental results obtained by the IG-based feature extraction method are still better than the MI-based method.

(3) Compared with the classification results of the Reuters-21578 dataset, the classification results obtained on the 20-Newsgroups dataset are slightly worse. This shows that the experimental results obtained by the same classification method on different text datasets are not the same.

5. Conclusion

With the development of internationalization, the demand for literary translation has increased dramatically. In order to make better use of these literary translation texts, this paper studies the classification of literary translation texts. Because the traditional classification method only pays attention to the words and phrases of the composition, and only knows which words correspond to which application fields, words are relatively isolated from each other and do not consider other characteristics such as word collocation, style, and typesetting. Therefore, this paper proposes a literary translation text classification method based on the distributed incremental sequence data mining algorithm. The method can fully mine the features of syntactic order and other features based on considering the characteristics of composition words and phrasing in different application fields, to discover more effective information, so it can strengthen the text classification effect. The experimental results also support this. This method also has some shortcomings, such as the classification accuracy is still far from being used in real production scenarios, and the classification accuracy needs to be further improved. The future research will be mainly carried out from the following two aspects: First, we will continue to deeply study the characteristics of composition words and wording in different application fields and try to compile a relatively high-quality stop word list, to improve the efficiency of the classification of literary translation texts by this method. The second is to introduce more feature extraction methods to improve the classification performance. Because a good feature extraction method can greatly improve the text classification performance, which is also shown in the experiments.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9857513, 1 page
https://doi.org/10.1155/2023/9857513

Retraction

Retracted: Research on English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Yi, "Research on English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation," *Security and Communication Networks*, vol. 2022, Article ID 1865451, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1865451, 12 pages
https://doi.org/10.1155/2022/1865451

Research Article

Research on English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation

Yana Yi

Henan Institute of Technology, Xinxiang, 453000, China

Correspondence should be addressed to Yana Yi; 13673523391@hait.edu.cn

Received 21 February 2022; Revised 15 March 2022; Accepted 28 March 2022; Published 4 May 2022

Academic Editor: Chin-Ling Chen

In order to improve the evaluation effect of English teaching reading quality, this paper applies the cognitive diagnosis evaluation algorithm to the teaching reading quality evaluation process. Moreover, this paper compares the verification effects of *Q*-matrix from the perspective of nonparameterization and parameterization and considers their performance through simulation research. In addition, this paper applies the method to multilevel scoring empirical data to verify the effectiveness of the cognitive diagnostic evaluation algorithm. Finally, this paper combines cognitive diagnosis and evaluation algorithms to construct an English teaching reading quality evaluation system. The experimental research results show that the English teaching reading quality evaluation method based on cognitive diagnostic evaluation proposed in this paper can effectively improve the reading quality of English teaching.

1. Introduction

In the English adaptive reading system, we use the length of the reading material, sentence structure, article type, vocabulary phrase situation within the concept, and other language ability characteristics to characterize the difficulty of reading. If we use people to obtain feature variables manually, this will be a very ambitious task. However, the current level of information technology is still unable to automatically obtain these characteristics. Therefore, the function of this link is to obtain the language feature quantity of the reading material as automatically as possible and manually label the remaining feature quantity [1].

In the field of natural language processing, machine reading comprehension tasks play an important role. Humans develop their understanding of texts by doing reading comprehension as they grow up. Machine reading comprehension is to let the machine imitate the behavior of human reading comprehension and improve the machine's comprehension ability by reading a given article to the machine and answering relevant questions. Early machine reading comprehension mainly extracted features manually, designed task rules, and then used conventional machine learning classification algorithms to make predictions. This early machine reading comprehension system can be thought of as an expert system that can answer questions in a particular domain according to a set of logical rules derived from professional knowledge. Because it only focuses on a certain field, the design is simple and easy to implement, but the application is narrow, lacks common sense, and has poor versatility, which cannot meet the needs of current practical applications. With the continuous improvement of the quality of data sets, the rapid expansion of scale, and the emergence of deep neural networks, machine reading comprehension has developed rapidly. At this stage, machine reading comprehension automatically learns task-related features and rules through deep learning networks and uses various attention mechanisms to obtain interactive information, which greatly improves the performance of the model.

Computer-assisted instruction is abbreviated as CAI, which is an automated computer-aided instruction technology that has developed rapidly in the past 30 years. CAI integrates many disciplines, such as computer science, pedagogy, psychology, and electronic education. Moreover, it has many advantages, such as the scientific nature of knowledge combination, the intuitiveness of expression, the efficiency of classroom teaching, the timeliness of feedback correction, the initiative of learning activities, the simulation of process demonstration, the scalability and reproducibility, and so on. Therefore, it shows great vitality [2]. Carrying out computer-assisted teaching is not only in keeping with the trend of world education modernization, but also the needs of China's education reform and training of talents. At present, there are many forms of computer-assisted English teaching, such as helping teachers prepare lessons, select questions, write papers, read, test, and help students fill in exercises, memorize words, and correct mistakes. However, most of them are simple, fixed, and unchanging models, which cannot be different from person to person; that is, there is no intelligence [3].

This article combines cognitive diagnostic evaluation methods to evaluate and analyze the quality of English teaching reading, build an intelligent system, and verify the performance of the system to improve the level of English reading teaching.

Organizational structure of this paper is as follows: The first part studies the current situation of reading in English teaching, puts forward relevant problems, and lays the foundation for the creation of this paper. The second part analyzes the current situation of mobile reading, summarizes related technologies, and leads to the research content of this paper. The third part analyzes the main algorithmic cognitive diagnostic evaluation methods in this paper, which provides the algorithm basis for intelligent reading quality evaluation. The fourth part constructs an English teaching reading quality evaluation system based on cognitive diagnostic evaluation and analyzes the effect of the system. Finally, the conclusions of this paper are summarized in the conclusion section.

The main contribution of this paper is to compare the validation effects of Q-matrices from nonparametric and parametric perspectives, consider their performance through simulation studies, and apply the method to empirical data on multilevel scoring to verify that the method works in empirical studies. The performance in the data provides an effective method for intelligent system data processing.

2. Related Work

When studying the development status of mobile reading, the author also combed the existing literature. Literature [4] elaborated on the understanding of the concept of mobile reading behavior and carried out a quantitative analysis of mobile reading behavior, mainly through questionnaire surveys. It includes the purpose of the user's mobile reading, reading context, reading frequency, and main reading content. Literature [5] provides some inspiration for related research on the understanding of mobile reading, such as personalized and social reading. In the research that affects the efficiency of mobile reading, literature [6] studies the information presentation problem of mobile terminals from the aspect of reading satisfaction with mobile reading. Literature [7] focuses on the usability and ease of use of reading products in the bookstore. The point of view in the article is that information presentation is an important part of the construction of a good reading experience because information acquisition is the essence of reading. In mobile reading, it is very important to present information in a way suitable for mobile reading. Literature [7] took context analysis and related theories as the starting point, conducted an in-depth analysis of the group's mobile reading behavior, and finally extracted ten main elements that affect the user experience of mobile reading devices.

The task of machine reading comprehension is based on a given article and article-related questions, modeling the article and article-related questions by constructing a suitable model, and selecting or generating answers to the questions. It uses triples (C, Q, and A) to describe the problem, where C represents the article, Q represents the article-related question, and A represents the answer to the question [8]. Machine reading comprehension can be divided into four types: closed type, multiple-choice type, text extraction type, and answer generation type. Close machine reading comprehension is based on contextual information to predict the words that are dug out in the sentence of the article, and most of the words that are dug out are entity words. The characteristic of multiple-choice machine reading comprehension is that after the machine understands a given article, it starts from related questions and selects a correct answer from the set of candidate answers [9]. Text extraction machine reading comprehension requires the machine to extract a certain piece of continuous text from a given article as the answer to the question. It should be noted that the answer may be a word or a sequence of multiple consecutive words [10]. Compared with a single entity of the close class, the task of text extraction is more demanding and more challenging. Answer generation machine reading comprehension is the most flexible reading comprehension task. It has a variety of answers. It may require a model to extract a certain fragment from the original text as an answer like text extraction. It may also require a model to understand the article. After summary, generate the answer yourself (the original text does not) [11]. The answer generation category is closer to human answering questions, involving multiple rounds of question and answer, multihop reasoning, and other technologies. The BiDAF model published in [12] is mainly for experiments on the CNN/DM data set, which belongs to the close-filled machine reading comprehension, and has been innovated in the coding layer and the information interaction layer. The coding layer of BiDAF is formed by stacking multiple subcoding layers, using vector representations of various granularities. First, it maps each character of the input word into a vector and uses the CNN network and Max-pooling method to combine all the character vectors of each word into a character-level vector representation. Then, use the pretrained word vector GloVe to query the word-level vector representation of each word. Finally, the character-level vector representation and word-level vector representation of each word are spliced before and after to form a vector representation with a larger dimension, and then input to the two-way LSTM network to obtain the same dimension of each word, which contains certain context information The vector representation. Literature [13] proposed a new two-way attention flow mechanism. This two-way attention mechanism calculates query-to-context (Q2C) attention and context-2-query (C2Q) attention and obtains mutual information such as problem-perceived article representation and article-perceived problem representation. The R-Net introduced in [14] belongs to the text extraction machine reading comprehension. It also uses GloVe's character-level and word-level embedding representations in the programming layer to be stitched together in a front-to-back manner. At the information interaction layer, a gated attention mechanism is used for the article sequence and question sequence to extract the problem-aware article representation, and then the same gated attention mechanism is used to self-match the problem-aware article representation obtained in the previous step so that the problem-aware article means obtaining the overall article information [15]. In the answer output layer, the pointer network is used to output the location information of the beginning and end of the answer. The QANet proposed in [16] is also a text extraction machine reading comprehension, which uses a convolutional neural network (CNN) to replace the traditional recurrent neural network (RNN), shortening the training time. QANet is mainly formed by stacking multiple blocks in the coding layer. Each block is composed of a convolutional network layer (convolution layer) plus a self-attention layer plus a fully connected network layer (feedforward layer) composition. The SDNet submitted in [17] belongs to the answer generation machine reading comprehension. It uses a pretrained language model to represent BERT to improve the coding layer and uses FusionNet's ideas for multiple fusions in the information interaction layer. Specifically, SDNet combines the GloVe word-level vector representation and the fixed weight and weighted recombined BERT vector representation at the coding layer to obtain a new vector representation. Residual connection is used multiple times in the information interaction layer; that is, the input of the previous layer and the output of the previous layer are combined into the input of the next layer.

3. Cognitive Diagnosis and Evaluation Methods

In this paper, the cognitive diagnostic evaluation algorithm is used as the English teaching reading quality evaluation algorithm to provide a reference for the construction of the subsequent English teaching reading quality evaluation system, and then the algorithm is studied.

3.1. Matrix Verification Metrics. In this study, the nonparametric Q-matrix verification method is expanded, and the corresponding Q-matrix verification index is extended to the multilevel scoring problem, and the corresponding method is denoted as R^P. Based on the q vector of a certain item in the test, if the sum of the residual squares RSS between the observation response and the ideal response on this item can be minimized, it indicates that the q vector of the item is correctly specified. The calculation formula of this method under multilevel scoring can be expressed as [18]

$$R_j^P = \sum_{i=1}^{N} \left(u_{ij} - E\left(U_{ij}|q_j, \widehat{\alpha}_i\right)\right)^2. \tag{1}$$

Among them, $E(U_{ij}|q_j)$ represents when the attribute vector of the jth question is q_j, the expected value of the ith subject's score on this question. The superscript P in R_j^P represents the multilevel scoring, which is to distinguish it from the statistics in the second-level scoring, and $\widehat{\alpha}_i$ is the estimated value of the participant's attribute mastery mode. Further, if the statistics are classified according to the knowledge attribute vector α of the subjects, then formula 1 can be transformed into the following form:

$$R_j^P = \sum_{m=1}^{2^K} \sum_{i\in C_m} \left[u_{ij} - E\left(U_{jm}|q_j, \widehat{\alpha}_i\right)\right]^2. \tag{2}$$

Among them, C_m is the first potential mastering category of m, where $E(U_{jm}|q_j, \widehat{\alpha}_i) = \sum_{h=1}^{H} h \cdot P(Y_t = h|\widehat{\theta})$. If considering the different distributions of each attribute mastering mode, formula 2 can be further transformed into the following formula [19]:

$$R_j^P = \sum_{m=1}^{2^K} \pi_{\alpha_m} \sum_{i\in C_m} \left[u_{ij} - E\left(U_{jm}|q_j, \widehat{\alpha}_i\right)\right]^2, \tag{3}$$

π_{α_m} is the posterior probability distribution of C_m, and formula 3 can be understood as the sum of the expected residual squares between the observation score and the expected score.

For the subject's classification problem, if the $\eta_i(\widehat{\alpha}_i, q_j)$ corresponding to the knowledge attribute mastering mode $\widehat{\alpha}_i$ can minimize $\mathrm{d}_{wh}(u_i, \eta_i(\widehat{\alpha}_i, q_j))$, then $\widehat{\alpha}_i$ is the estimated value of the subject's knowledge attribute mastering mode. The specific formula is as follows [20]:

$$\mathrm{d}_{wh}\left(u_i, \eta_i\left(\widehat{\alpha}_i, q_j\right)\right) = \sum_{j=1}^{J} \frac{1}{\overline{p}_j\left(1 - \overline{p}_j\right)} \left|u_{ij} - \eta_i\left(\widehat{\alpha}_i, q_j\right)\right|. \tag{4}$$

In the above formula, h represents the hth ideal attribute mastering model, $1/\overline{p}_j(1 - \overline{p}_j)$ is the weighted part of the question, and $\overline{p}_j$ represents the proportion of subjects who answered the question j correctly or the probability of answering the question correctly. $\overline{p}_j(1 - \overline{p}_j)$ is the variance of the participant's observation and answer on question j. Therefore, this indicator is more inclined to choose topics with small variance. Through the previous analysis, a two-step iterative algorithm can be used to estimate the attribute vector of the question based on the R_j^P statistic. The first step is to use the weighted Hamming distance $\mathrm{d}_{wh}(u_i, \eta_i(\widehat{\alpha}_i, q_j))$ to classify the subjects, and the second step is to use R_j^P to estimate the attribute vector of the subject based on the classification of the subjects.

3.2. Algorithmic Simulation Assumptions. This study simulates the data under the conditions of different number of subjects (400, 600, 800, and 1000) and different error q

vectors (5%, 10%, and 15%). There are a total of $4 \times 3 = 12$ experimental conditions, each of which is repeated 100 times. In this study, R software was used to generate simulation data.

(1) *Mock Test Q-Matrix*

For the Q-matrix, the knowledge attributes required for an answer category refer to the attributes required by the candidate to correctly reach the category after completing all the previous steps.

(2) *Simulation of Q-matrix Containing Errors*

The initial Q-matrix Q_0 is constructed on the basis of the real Q-matrix, and the questions with incorrect calibration in Q_0 are randomly selected according to three proportions (5%, 10%, and 15%, resp.). The selected wrong item is randomly selected among $2^K - 2$ possibilities (cannot be a vector of all 0s and a correct vector).

(3) *Simulate the Knowledge State of the Subjects*

It is assumed that the knowledge state of the subjects obeys a uniform distribution; that is, the number of subjects under each knowledge attribute mastery mode is similar.

(4) *Simulation Question Parameters*

The simulation of the question parameters is completed according to the following rules; that is, the highest category is $S_j(h|\alpha^*_{ljh} = 1) = 0.9$, and the lowest category is $S_j(h|\alpha^*_{ljh} = 0) = 0.1$. When there are more than two answer categories, the middle category probability is randomly selected from the uniform distribution $\mathrm{U}(S_j(h|\alpha^*_{ljh} = 0), S_j\ (h|\ \alpha^*_{ljh} = 1))$, and it is ensured that the more attributes are mastered, the greater the category probability is.

(5) *Simulate the Subject's Answer*

After simulating the student's knowledge state, Q-matrix and item parameter values in the above steps, the response probability of the participant on the item is calculated. Participants' response scores were simulated based on the item response function of the sequential GDINA model.

(6) In order to evaluate the performance of the R^P method in the Q-matrix estimation, the indicators at the Q-matrix level are, respectively, used to evaluate the estimation accuracy, which includes the number of successful estimations, the evaluation indicators of the item level, the number of criteria for the item mode, the evaluation index of the attribute level, and the average criterion number of the item attributes. In terms of estimating efficiency, the average number of iterations and the average running time are used to evaluate.

Among them, the number of successful estimates N_{success} is as follows: it represents the number of times the Q-matrix is completely correctly estimated in 100 batches of randomly generated data.

$$N_{\text{success}} = \sum_{r=1}^{100} \mathbf{I}\left(Q^{PT} == \widehat{Q}_r^{P}\right). \tag{5}$$

Among them, the indicator function $\mathbf{I}(Q^{PT} == \widehat{Q}_r^{P})$ indicates whether the rth estimated Q -matrix is exactly the same as the real Q-matrix, that is, whether the Q matrix containing errors is successfully corrected. $\mathrm{I}(Q^{PT} == \widehat{Q}_r^{P}) = 1$ means successful estimation; otherwise, it is 0.

The criterion number of question patterns (PMN) is as follows: the algorithm calculates the consistency (Pattern Match Number, PMN) between the measurement pattern of all questions in the Q-matrix estimated each time and the measurement pattern of the real Q-matrix question (Pattern Match Number, PMN) and calculates the average of 100 experiments.

$$\mathrm{PMN} = \frac{\sum_{r=1}^{100}\sum_{j=1}^{J} \mathrm{I}\left(q_j == \widehat{q}_j\right)}{r}, \tag{6}$$

where $\mathbf{I}(q_j == \widehat{q}_j)$ indicates whether the attribute vector of item j in the r batch of data is correctly estimated. 1 indicates the correct estimation; otherwise, it is 0.

Item attribute average criterion number (ARN) is as follows: the algorithm calculates the average correct recovery amount of item attributes in 100 repeated experiments, which reflects the probability of knowledge attribute recovery (Attribute Recovery Number, ARN).

$$\mathrm{ARN} = \frac{\sum_{r=1}^{100}\sum_{j=1}^{J} \mathrm{I}\left(q_{jk} == \widehat{q}_{jk}\right)}{100 \times J}. \tag{7}$$

When the Q-matrix is not successfully restored, the item model criterion number PMN and the item attribute average criterion number ARN describe the extent to which the estimation method restores the item attribute vector. The higher the PMN and ARN, the more accurate the estimation of the method.

The average number of iterations (AIN) is as follows: the algorithm calculates the average number of iterations in 100 repeated experiments (Average Iterative Number).

$$\mathrm{AIN} = \frac{\sum_{r=1}^{100} It\mathrm{Num}_r}{100}, \tag{8}$$

where $It\mathrm{Num}_r$ refers to the number of iterations used in the rth estimation.

Average running time (ART) is as follows: the algorithm calculates the average running time ART (Average Running Time) of the Q-matrix verification process in 100 repeated experiments.

$$\mathrm{ART} = \frac{\sum_{r=1}^{100} t_r}{100}, \tag{9}$$

t_r refers to the running time used in the rth experiment. Analyzing the same batch of data, the smaller ART indicates the higher efficiency of the algorithm.

Figure 1 and Figure 2 show the changes in the number of successful estimates of the Q-matrix under different types of grouping. Figure 1 is grouped according to different numbers of subjects, and Figure 2 is grouped according to different Q-matrix error ratios.

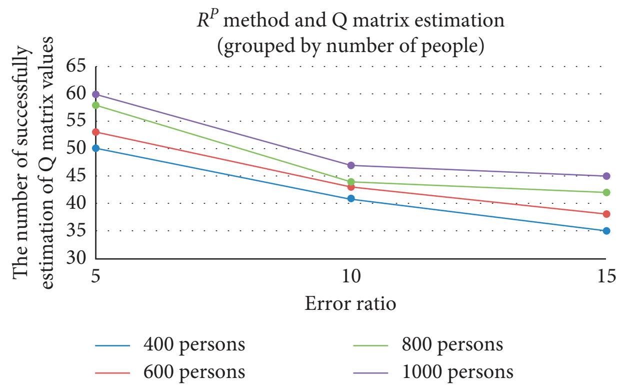


Figure 1: The number of Q-matrices successfully estimated by the R^P method (grouped by the number of people), range [0, 100].

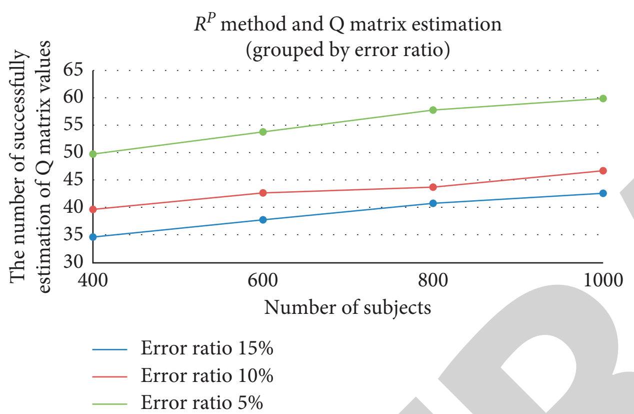


Figure 2: The number of Q-matrices successfully estimated by the RP method (grouped by the error ratio of the elements in the Q-matrix), range [0, 100].

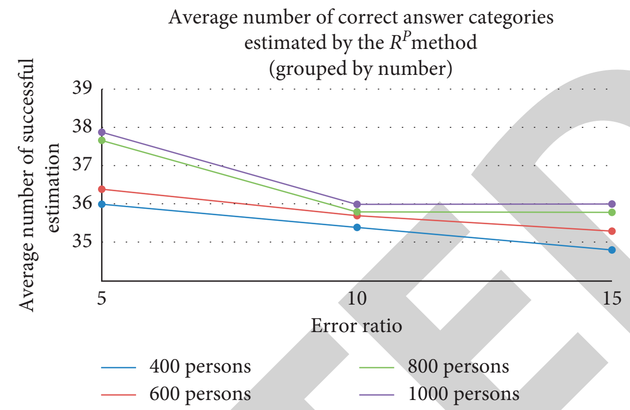


Figure 3: The average number of answer categories (grouped by the number of people) of questions successfully estimated by the R^P method, range [0, 39].

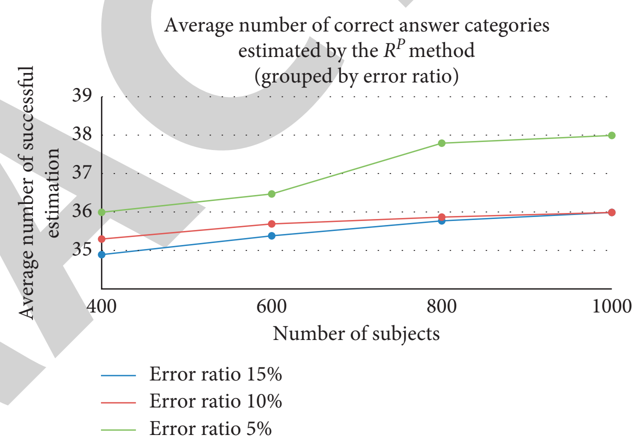


Figure 4: The average number of answer categories (grouped according to the error ratio) of successfully estimated questions by method R^P, range [0, 39].

As shown in Figure 1, it describes the trend of estimating the Q-matrix with the number of people using the R^P method. It is easy to see the influence of the increase in the number of people and the decrease in the error rate on the Q-matrix estimation effect under the nonparametric method. Obviously, the larger the number of people and the lower the error rate, the better the Q-matrix estimation effect.

In order to further verify the performance of the RP method in estimating the Q-matrix, this study examined the estimation success rate of the method step by step.

Figures 3 and 4, respectively, show the estimated success numbers of the attribute vectors corresponding to the correct answer categories under the conditions of grouping by the number of people and grouping by the error ratio when using the R^P method to estimate the Q-matrix. The number of answer categories in the entire Q-matrix is 39. Therefore, the closer this indicator is to 39, the more accurate the estimate is.

The basic idea of the Q-matrix estimation algorithm is as follows: when the number of subjects is large enough, we analyze the test data based on the correctly defined Q-matrix to obtain model parameters and subject classification parameters. The expected response distribution and the observed response distribution obtained by calculation should be equal, which can be represented by the following pseudocode:

$$\text{if } Q^{P'} = Q^{P}_{\text{true}}, \tag{10}$$

$$\text{then } P\left(R|Q^{P'}, p\right) = \hat{P}(R). \tag{11}$$

Among them, Q^{P_1} is the Q-matrix (also called the estimated value of the matrix under the alternative multilevel scoring diagnostic test), Q^{P}_{true} is its true value, $P(R|\,Q^{P'}, p)$ represents the distribution of the answer vector R determined by the model parameters and the overall distribution, and $\hat{P}(R)$ represents the observed distribution of the answer vector R. At the same time, there are

$$P\left(R|Q^{P'}, p\right) = \sum_{\alpha} p_{\alpha} \prod_{j=1}^{J} P\left(R^{j}|Q^{P'}, \alpha\right),$$
$$\widehat{P}(R) = \frac{1}{N}\sum_{i=1}^{N} I\left(R_i = R\right). \tag{12}$$

Among them, p_α represents the distribution of the attribute mastering model α in the population, and R_i represents the response vector of the subject i. A key concept of the S^P method is the S^P matrix (T-matrix); its function is to describe the expected response distribution.

(1) For a single question (we assume that the highest score of the question is H; that is, the score interval is $[0, H]$, and there is a total of $H + 1$ answer categories), then there are

$$\begin{aligned} P\left(R^{j} = 1|Q^{P'}, p\right) &= \sum_{\alpha} p_{\alpha} P\left(R^{j} = 1|Q^{P'}, \alpha\right) \\ &= B_{Q^{P'}}(j)p, \\ P\left(R^{j} = H|Q^{P'}, p\right) &= \sum_{\alpha} p_{\alpha} P\left(R^{j} = H|Q^{P'}, \alpha\right) \\ &= B_{Q^{P'}}(j)p. \end{aligned} \tag{13}$$

(2) For item pairs (here for convenience, we assume that each item has the same answer category. In practical applications, the answer category of each item may be different), then there are

$$\begin{aligned} & P\left(R^{j1} = 1, R^{j2} = 1|Q^{P'}, p\right) \\ & = \sum_{\alpha} p_{\alpha} P\left(R^{j1} = x_{j1}|Q^{P'}, \alpha\right) P\left(R^{j2} = x_{j2}|Q^{P'}, \alpha\right) \\ & = B_{Q^{P'}}(j1, j2)p, \\ & P\left(R^{j1} = H, R^{j2} = H|Q^{P'}, p\right) \\ & = \sum_{\alpha} p_{\alpha} P\left(R^{j1} = H|Q^{P'}, \alpha\right) P\left(R^{j2} = H|Q^{P'}, \alpha\right) \\ & = B_{Q^{P}}(j1, j2)p. \end{aligned} \tag{14}$$

Therefore, there are a total of C_J^2 question pairs. In the same way, there are three combinations of topics, up to the combination of J topics. In this way, the T-matrix of size $(2^{JH} - 1) \times 2^K$ can be constructed:

$$T = \begin{array}{c} \\ 1 \\ 2 \\ \cdots \\ j \\ 1u2 \\ \cdots \\ \end{array} \begin{bmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_{2^K} \\ P_{\boldsymbol{\alpha}_1,1} & P_{\boldsymbol{\alpha}_2,1} & \cdots & P_{\boldsymbol{\alpha}_2 K,1} \\ P_{\boldsymbol{\alpha}_1,2} & P_{\boldsymbol{\alpha}_2,2} & \cdots & P_{\boldsymbol{\alpha}_2 K,2} \\ \vdots & \vdots & \vdots & \vdots \\ P_{\boldsymbol{\alpha}_1,J} & P_{\boldsymbol{\alpha}_1,J} & \cdots & P_{\boldsymbol{\alpha}_2 K,J} \\ P_{\boldsymbol{\alpha}_1,1\cup 2} & P_{\boldsymbol{\alpha}_2,1\cup 2} & \cdots & P_{\boldsymbol{\alpha}_2 K,1U2} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{15}$$

We assume that the row vectors of the T matrix are $B_{Q^{P'}(1)}, \ldots\ldots, B_{Q^{P'}(1,2)}$, respectively; then formula 9 can be expressed as

$$T_{\left(Q^{P'}\right)} = \begin{pmatrix} B_{Q^{P'}(1)} \\ \ldots\ldots \\ B_{Q^{P'}(J)} \\ B_{Q^{P'}(1,2)} \\ \ldots\ldots \end{pmatrix}. \tag{16}$$

Further, according to formula 5 to formula 8, the T-matrix can be expressed as

$$T_{\left(Q^{P'}\right)} = \begin{pmatrix} P\left(R^{1} = 1|Q^{P'}, p\right) \\ \ldots\ldots \\ P\left(R^{J} = 1|Q^{P'}, p\right) \\ P\left(R^{1} = 1, R^{2} = 1|Q^{P'}, p\right) \\ \ldots\ldots \end{pmatrix}. \tag{17}$$

The β vector is another important concept of this method, which is the column vector corresponding to equation 10. The score vector is the ratio of the number of people answering the question combination, and it represents the distribution of observation scores. When $N \longrightarrow \infty$, if all parameters are calibrated correctly, according to the law of large numbers, there is

$$\beta = T_{\left(Q^{P'}\right)} p. \tag{18}$$

Therefore, the objective function of the matrix verification under the multilevel scoring diagnostic test can also be expressed as [21]

$$S_{p\left(Q^{P'}\right)} = \left| T_{\left(Q^{P'}\right)} p - \beta \right|. \tag{19}$$

Among them, |.....| is the Euclidean distance. Since the parameters are unknown, it is necessary to estimate each parameter based on the model and use maximum likelihood estimation to estimate each parameter.

This research is based on the following assumptions. We assume that a Q-matrix has been defined by experts in the relevant knowledge domain, that is, the initial Q-matrix. Moreover, this initial Q-matrix contains only partial errors (i.e., some of the elements are incorrectly defined, and the remaining elements are correctly defined).

This paper simulates the data under different number of subjects (800, 1000, 2000, and 4000) and different error attribute ratios (5%, 10%, and 15%). There are a total of $4 \times 3 = 12$ experimental conditions, and simulations are performed 100 times under each condition. In this study, R software was used to generate simulation data.

(1) Test the truth value of the Q-matrix.

The algorithm uses restricted Q-matrix, which is denoted as Q^{PT}.

(2) The algorithm simulates the initial Q-matrix.

Based on the real Q-matrix, the algorithm simulates the initial Q-matrix containing a certain proportion of errors, denoted as Q_0^P, and Q_0^P contains the incorrectly calibrated q vector. The wrong questions are randomly selected according to a certain percentage (5%, 10%, and 15%, resp.), and the selected wrong items are randomly selected among the possibilities in $2^K - 2$.

(3) Simulate the participant's knowledge state.

The knowledge state of the subjects is simulated in a uniform distribution; that is, the number of subjects under each knowledge attribute mastery mode is similar.

(4) Simulate question parameters.

The simulation of the topic parameters is completed according to the following rules; that is, the highest category is $S_j(h|\alpha_{ljh}^* = 1) = 0.9$, and the lowest category is $S_j(h|\alpha_{ijh}^* = 0) = 0.1$. When there are more than two answer categories, the middle category probability is randomly selected from the uniform distribution $\mathrm{U}(S_j(h|\alpha_{ljh}^* = 0), S_j(h|\alpha_{ljh}^* = 1))$, and it is ensured that the more attributes are mastered, the greater the category probability is.

(5) Simulate the subject's answer.

After simulating the student's knowledge state, Q-matrix, and item parameter values in the above steps, the response probability of the participant on the item is calculated. Participants' response scores were simulated based on the item response function of the sequential GDINA model.

$$\mathrm{P}\left(X_j = \mathrm{h}|\alpha_c\right) = \left[\left(1 - S_j\left(h + 1|\alpha_c\right)\right)\right] \prod_{x=0}^{h} S_j\left(x|\alpha_c\right). \qquad (20)$$

Among them, $S_j(x|\alpha_c)$ is the probability that the participant scores at step x on question j. It can use common cognitive diagnosis model functions, such as DINA or GDINA.

(6) Evaluation index

The evaluation indicators used in this study are the same as those in study one.

The algorithm uses the S^P method to estimate the results of the Q-matrix of the multilevel scoring diagnostic test. This research mainly focuses on the effect of Q-matrix estimation when the sample is large. On the whole, the performance of the parameterized S^P method is when the test sample is large. In most cases, the estimated success rate of the entire Q-matrix exceeds 50% [22].

It can be further seen from the results that as the sample increases, the success rate of the S^P method in estimating the Q-matrix will increase. As the correct elements in the Q-matrix increase, that is, the error rate is reduced, the success rate of the S^P method in estimating the Q-matrix will also increase.

Figures 5 and 6 show the changes in the number of successful estimates of the Q-matrix in the case of grouping

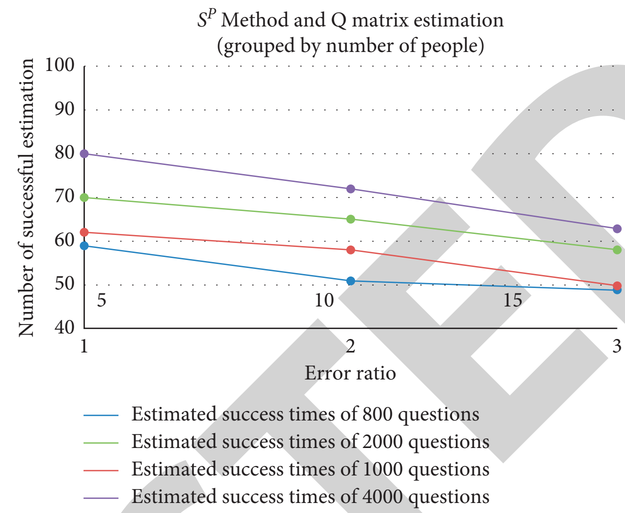


FIGURE 5: The number of Q-matrices successfully estimated by the S^P method (grouped by the number of people), range [0, 100].

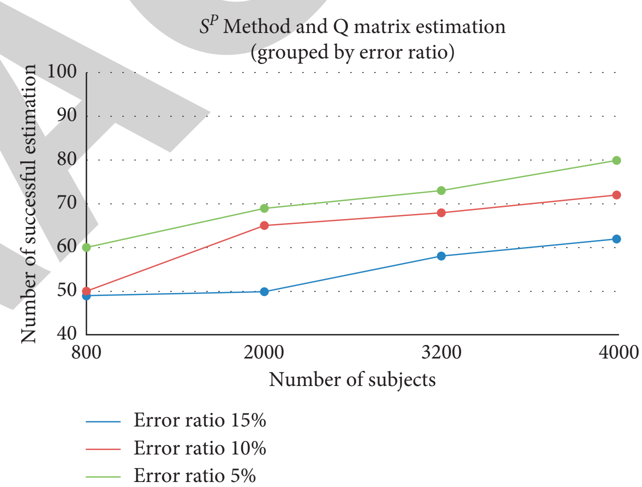


FIGURE 6: The number of Q-matrices successfully estimated by the S^P method (grouped by the error ratio of the elements in the Q matrix), range [0, 100].

by different types. Figure 5 is grouped according to the number of participants, and Figure 6 is grouped according to the Q-matrix containing different error ratios.

Figure 5 describes the change trend of the Q-matrix estimated by the S^P method with the number of people. It is easy to see the influence of the increase in the number of people on the parameterization method. Under different error ratio conditions, as the sample size increases, the effect of Q-matrix estimation gradually improves. When the error ratio of Q-matrix is 10% and 15%, and the number of subjects is 800, the correct estimation rate of Q-matrix is slightly higher than 50%. However, when the error rate is 5% and the number of subjects is different, the number of successful estimates of the Q-matrix is significantly higher than the error rate of 10% and 15%. Moreover, under the condition of a sample size of 4000, the success rate of Q-

matrix estimation reaches 80%. From Figure 6, it is easy to see the influence of the increase in the number of people on the estimation effect of the parameterization method. Under the condition of different error ratios, the increase in the number of people is still more obvious for the improvement of the estimated success rate, which is different from the R^P method.

In order to further evaluate the performance of the S^P method in estimating the Q-matrix, we examined the estimated success rate of the method step by step.

For the change trend of the successful estimation of the average category attribute vector in the entire Q-matrix, we refer to Figures 6 and7. The average Q-matrix element estimates the correct number, referred to in Figure 8.

It can be seen that on the one hand, as the proportion of error elements in the Q-matrix increases, whether it is the estimated success rate of the entire Q-matrix, the average estimated success rate of the answer category attribute vector in the Q-matrix, or the average estimated success rate of the elements in the Q-matrix will have a corresponding decline. On the other hand, as the number of subjects increases, the estimation accuracy indicators will increase to varying degrees.

It can be seen from Figure 6 that the estimated success times of the average correct answer category attribute vector increases as the proportion of errors decreases. When the sample is 4000, the average number of categorical attribute vectors of estimated errors is less than one under different error ratio conditions. However, when the number of subjects is 800, 1000, and 2000, the average number of categorical attribute vectors estimated to be incorrect reaches 2 or more under the conditions of 10% and 15% error ratios. It can be seen from Figure 7 that as the sample size increases, the estimated success times of the average answer category attribute vector increase significantly.

It can be seen from Figures 8 and9 that when the error ratio contained in the Q-matrix is only 5%, 1000 test data can achieve a higher element estimation success rate. However, when the error rate reaches a higher 10% or 15%, the estimated success rate of the elements in the Q-matrix is not high when the number of subjects is 800, 1000, and 2000. Furthermore, when the sample size reaches 4000, a higher Q-matrix element estimation success rate can be obtained. Moreover, when the sample size is 4000, the success estimates of the Q-matrix maintain a high and similar level under different sample size conditions. In general, the S^P method will obtain a relatively stable Q-matrix element estimation success rate when the sample reaches 4000 subjects.

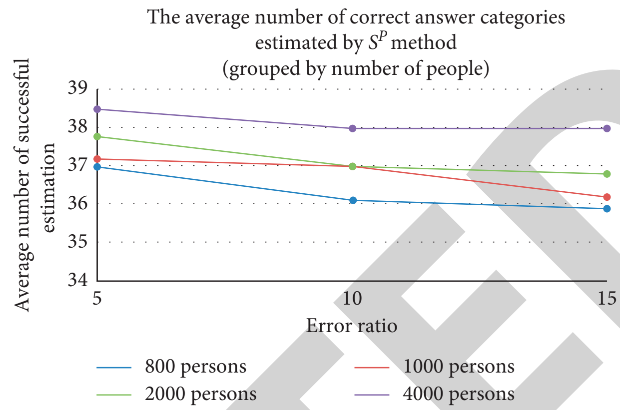


FIGURE 7: The average number of answer categories (grouped by the number of people) of successfully estimated questions by S^P method, range [0, 39].

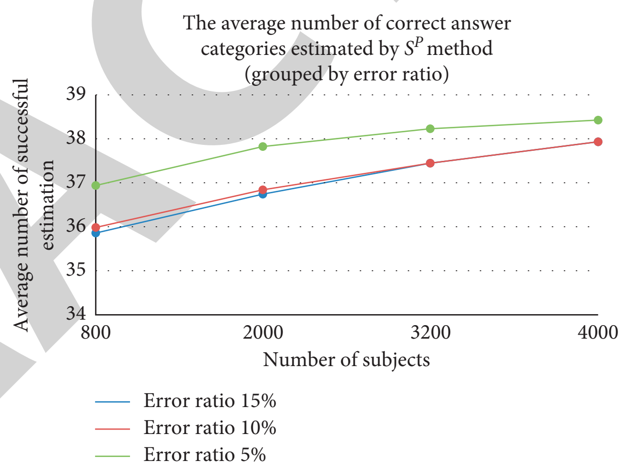


FIGURE 8: The average number of answer categories (grouped according to the error ratio) of successfully estimated questions by the S^P method, range [0, 39].

4. Research on the English Teaching Reading Quality Evaluation Method Based on Cognitive Diagnostic Evaluation

This article regards students' learning English as a structured activity to solve problems. In this activity, the problem solver starts from the starting state, goes through a series of intermediate states, and finally reaches the goal state. Artificial intelligence research shows that there are two ways to reach a goal. One is to give all possible states and test whether the final state meets the desired goal, and the other is to use the available additional information to find the correct solution path. In the process of learning or reading English, the text given is the initial state of the problem. The goal is to connect the content of the text with the reader's existing knowledge structure. Moreover, problem-solving activities form a problem space. In this problem space, according to the additional information that can be obtained, seek the correct solution path. This process is shown in Figure 10(a).

Figure 10(b) describes a very simple network used to analyze English sentences. This kind of network is called "S" network. In the network shown in Figure 10(b), from one state to another, the input signal must be confirmed as a verb. That is, one state must be connected to another state through necessary actions. Also, only the lead from the initial state is allowed, and there will always be a line starting from the initial state until the send line. The other is the NP network. The connection between the states is divided into a search line and a classification line, as shown in Figure 10(c).

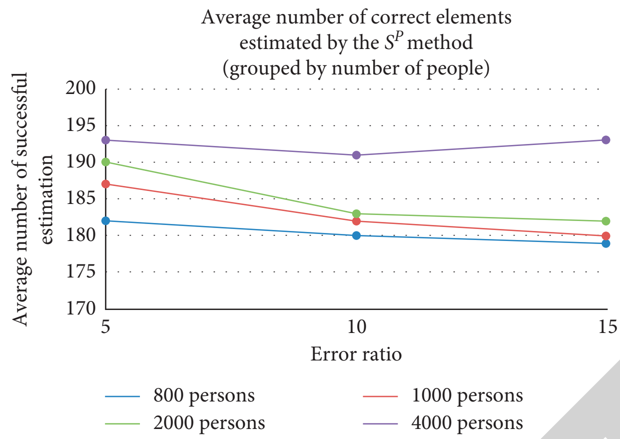


Figure 9: The number of attribute elements of the question successfully estimated by the S^P method, range [0, 195].

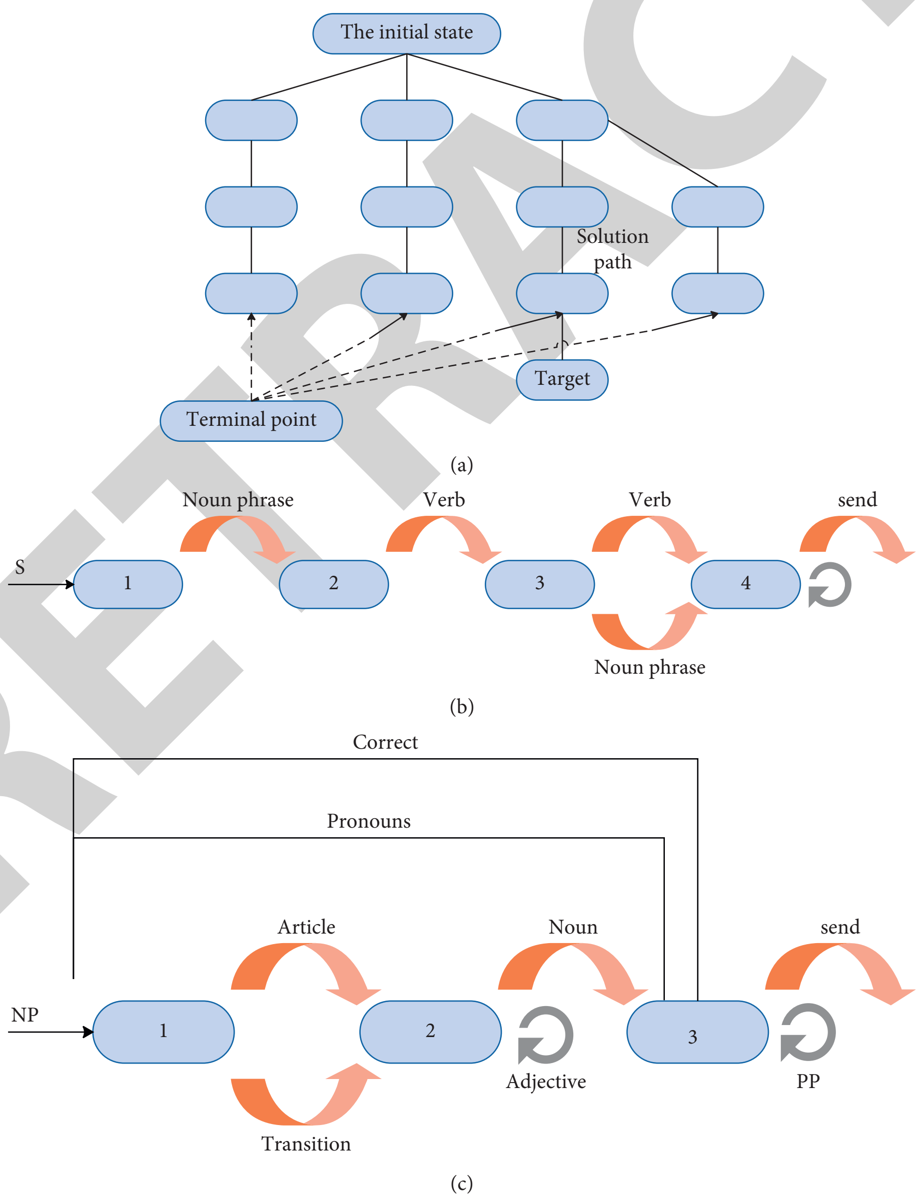


Figure 10: (a) Description of the problem space. (b) Simple transition network. (c) Simple NP network.

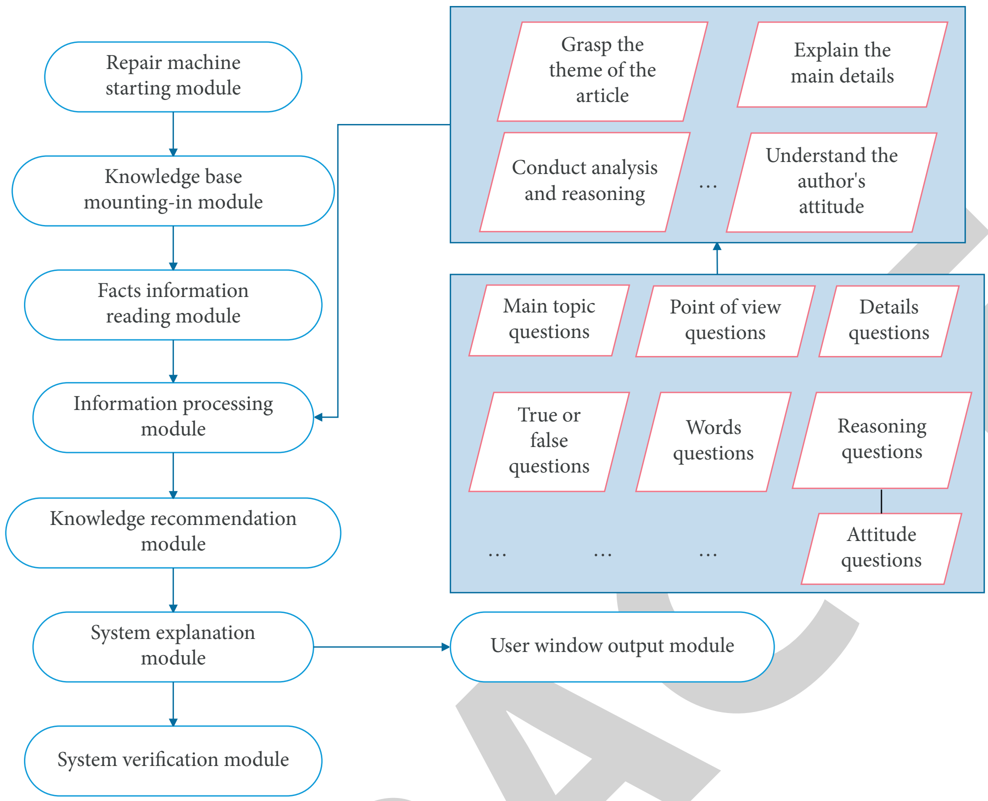


Figure 11: Reasoning network of the expert system for English reading teaching.

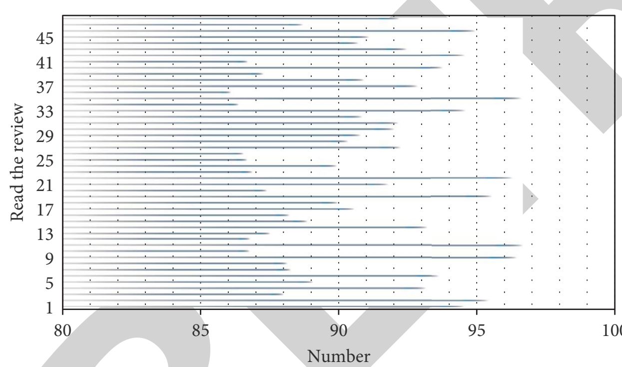


Figure 12: The effect of English teaching reading quality evaluation based on cognitive diagnostic evaluation.

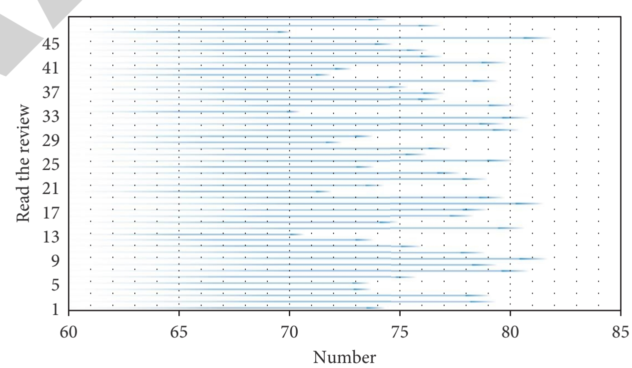


Figure 13: The effect of the English teaching reading quality evaluation system based on cognitive diagnostic evaluation on the improvement of teaching quality.

The design of the inference engine of the English Reading Teaching Expert System is based on fuzzy inference, reasoning combined with artificial neural network models, nonmonotonic reasoning, and mixed control strategies. The inference engine repeatedly matches the rules in the knowledge base against the conditions or known information in the field of English reading teaching. The reasoning network drawn in Figure 11 helps teachers and students understand the reasoning mechanism and interpretation mechanism of the expert system.

On the basis of the above research, this paper uses experimental teaching methods to study the English teaching reading quality evaluation system proposed in this paper and calculates the evaluation effect and teaching effect, as shown in Figure 12 and Figure 13.

From the above research, it can be seen that the English teaching reading quality evaluation method based on cognitive diagnostic assessment proposed in this paper can effectively improve the reading quality of English teaching.

5. Conclusion

The network platform can enable students to share resources. Moreover, it provides an appropriate amount of input every day to keep students in a dynamic language learning environment. Moreover, some students with higher learning ability and learning level can also share the high-quality resources they have obtained in extracurricular reading. In the open era, students have a wealth of life information, personalized life experience, and innovative learning methods. Moreover, in the process of cooperative learning, inquiry learning, and autonomous learning, they have also formed a rich and colorful curriculum resources between each other, which become the activity carrier of reading resources. In addition, the process of developing and using curriculum resources is the process of student learning. This article combines cognitive diagnostic evaluation methods to evaluate and analyze the English teaching reading quality, build an intelligent system, verify the performance of the system, and improve the level of English reading teaching. The experimental research results show that the English teaching reading quality evaluation method based on cognitive diagnostic evaluation proposed in this paper can effectively improve the reading quality of English teaching.

When estimating the *Q*-matrix in steps, under each condition, the two methods have higher success estimation rates in the first step, but with the increase of the answering steps, the two methods are more sensitive to the attributes corresponding to the later steps. The success rate of vector estimation will gradually decrease, which is also the main reason for the low success rate of estimation of the entire *Q*-matrix. The next step requires in-depth research on methods to improve the success rate of attribute vector estimation corresponding to the following steps.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This study was sponsored by Henan Institute of Technology, China.

References

[1] A. M. Songbatumis, “Challenges in teaching English faced by English teachers at MTsN Taliwang, Indonesia,” *Journal of foreign language teaching and learning*, vol. 2, no. 2, pp. 54–67, 2017.

[2] J. C. Richards, “Teaching English through English: proficiency, pedagogy and performance,” *RELC Journal*, vol. 48, no. 1, pp. 7–30, 2017.

[3] L. Susanty, Z. Hartati, R. Sholihin, A. Syahid, and F. Y. Liriwati, “Why English teaching truth on digital trends as an effort for effective learning and evaluation: opportunities and challenges: analysis of teaching English,” *Linguistics and Culture Review*, vol. 5, no. S1, pp. 303–316, 2021.

[4] A. S. Fatimah, S. Santiana, and Y. Saputra, “Digital comic: an innovation of using toondoo as media technology for teaching English short story,” *English Review: Journal of English Education*, vol. 7, no. 2, pp. 101–108, 2019.

[5] B. Ayçiçek and T. Yanpar Yelken, “The effect of flipped classroom model on students’ classroom engagement in teaching English,” *International Journal of Instruction*, vol. 11, no. 2, pp. 385–398, 2018.

[6] N. Guzachchova, “Zoom technology as an effective tool for distance learning in teaching English to medical students,” *Bulletin of Science and Practice*, vol. 6, no. 5, pp. 457–460, 2020.

[7] M. S. Hadi, “The use of song in teaching English for junior high school student,” *English Language in Focus (ELIF)*, vol. 1, no. 2, pp. 107–112, 2019.

[8] A. Mahboob, “Beyond global Englishes: teaching English as a dynamic language,” *RELC Journal*, vol. 49, no. 1, pp. 36–57, 2018.

[9] H. Sundari, “Classroom interaction in teaching English as foreign language at lower secondary schools in Indonesia,” *Advances in Language and Literary Studies*, vol. 8, no. 6, pp. 147–154, 2017.

[10] K. Kim and Y. Park, “A development and application of the teaching and learning model of artificial intelligence education for elementary students,” *Journal of The Korean Association of Information Education*, vol. 21, no. 1, pp. 139–149, 2017.

[11] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education–where are the educators?” *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–27, 2019.

[12] S. C. Yang, W. K. Vong, R. B. Sojitra, T. Folke, and P. Shafto, “Mitigating belief projection in explainable artificial intelligence via Bayesian teaching,” *Scientific Reports*, vol. 11, no. 1, p. 34, 2021.

[13] Y. Lee, “An analysis of the influence of block-type programming language-based artificial intelligence education on the learner’s attitude in artificial intelligence,” *Journal of The Korean Association of Information Education*, vol. 23, no. 2, pp. 189–196, 2019.

[14] J. M. Alonso, “Teaching explainable artificial intelligence to high school students,” *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 974–987, 2020.

[15] R. Yang, “Artificial intelligence-based strategies for improving the teaching effect of art major courses in colleges,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 22, pp. 146–160, 2020.

[16] P. Ribeiro, H. Simões, and M. Ferreira, “Teaching artificial intelligence and logic programming in a competitive environment,” *Informatics in Education*, vol. 8, no. 1, pp. 85–100, 2009.

[17] V. Rampton, M. Mittelman, and J. Goldhahn, “Implications of artificial intelligence for medical education,” *The Lancet Digital Health*, vol. 2, no. 3, pp. e111–e112, 2020.

[18] Q. Zhou, “Reforms in teaching the course of introduction to artificial intelligence,” *Solid State Technology*, vol. 64, no. 2, pp. 6331–6335, 2021.

[19] K. I. M. Haejoong and L. E. E. Sangmin, “Do less teaching, do more coaching: toward critical thinking for ethical applications of artificial intelligence,” *Journal of Learning and Teaching in Digital Age*, vol. 6, no. 2, pp. 97–100, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9853176, 1 page
https://doi.org/10.1155/2023/9853176

Retraction

Retracted: Application of Data Mining Algorithm in Agricultural Products Logistics Network Planning

Security and Communication Networks

Received 30 January 2023; Accepted 30 January 2023; Published 5 February 2023

Security and Communication Networks has retracted the article titled "Application of Data Mining Algorithm in Agricultural Products Logistics Network Planning" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] D.-Y. Yu and G.-L. Wu, "Application of Data Mining Algorithm in Agricultural Products Logistics Network Planning," *Security and Communication Networks*, vol. 2022, Article ID 5884050, 9 pages, 2022.
[2] L. Ferguson, "Advancing research integrity collaboratively and with vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5884050, 9 pages
https://doi.org/10.1155/2022/5884050

Research Article

Application of Data Mining Algorithm in Agricultural Products Logistics Network Planning

Dong-Yun Yu and Gui-Ling Wu

Xinyang Agriculture and Forestry University, Xinyang 464000, China

Correspondence should be addressed to Gui-Ling Wu; wgl@xyafu.edu.cn

Received 10 February 2022; Revised 8 March 2022; Accepted 26 March 2022; Published 4 May 2022

Academic Editor: Chin-Ling Chen

The location of agricultural products center is related to the overall operation efficiency of logistics system and has a great impact on the follow-up agricultural products distribution route planning. The distribution route planning directly determines the total transportation cost and customer service quality of the whole logistics system. In this study, the problem of location path under the background of agricultural products logistics is studied, and an algorithm of agricultural products logistics network planning based on space-time constraints is proposed. An algorithm based on K-means is used to cluster the time-space double factors of customers. Minimize the total distribution cost under the constraints of meeting customer needs and minimizing time window deviation. Experimental results show that the proposed algorithm has better performance through Taguchi analysis and master-slave two-layer model analysis, and its algorithm has better convergence and sensitivity and can better reduce the total cost of distribution and the length of transportation path.

1. Introduction

With the promotion of the rural revitalization strategy, the issues concerning agriculture, rural areas, and farmers have become fundamental issues concerning the national economy and people's livelihood [1]. China is a big agricultural country, and agriculture is the basic industry of the national economy. With the continuous improvement of residents' living standards, people's demand for fresh agricultural products is also increasing [2]. Agricultural products logistics distribution is an important link in the supply chain of fresh agricultural products, which determines the quality of the final fresh food. At present, China's fresh agricultural products logistics are still in the stage of development, and the annual transport process of rotting fruit and vegetables is worth up to 70 billion yuan, resulting in a huge economic loss [3]. Therefore, it is of great practical significance to accelerate the development of fresh agricultural products logistics distribution and optimize the location and path of agricultural products logistics [4].

The contributions of this study are listed as follows: (1) a k-means based clustering algorithm for customer spatial-temporal is designed. (2) Under the constraints of meeting customer requirements and minimizing time window deviation, the total cost is minimized.

2. Related Works

In order to cope with the rapidly growing demand for agricultural products, site selection and route planning of logistics outlets have become key links in supply chain management and logistics system construction [5]. It plays an important role in reducing the operation cost and improving the distribution efficiency and customer satisfaction. Logistics site selection is a medium and long-term strategy, as the hub of logistics activities directly affects the whole system planning [6]. It is interdependent with customer location and distribution route arrangement. If route planning is ignored in site selection, the cost of logistics distribution system will be increased. The specific route of vehicle distribution is a short-term operational decision, but the short-term decision will ultimately affect customer satisfaction and the total operating cost [7]. The actual distribution process is in a very complicated environment. In the process of distribution, it is easy to encounter all kinds of interference events, such as refrigeration unit cannot work as usual due to vehicle breakdown and traffic accidents caused by vehicles can not arrive on time. This will affect the execution of the initial distribution plan or even cause distribution disruption. At this time, it will not only affect the operation cost and increase the burden of enterprises but also aggravate the food decay and then affect the consumption safety of residents [8].

In recent years, scientific site selection and rational distribution path planning of agricultural products have attracted extensive attention of scholars at home and abroad. Literature [9] studied the VRP (vehicle routing problem) model of fruit and vegetable distribution with a time window and solved it by the tabu search algorithm. Literature [10] proposed the threshold acceptance algorithm for multi-model VRP of fresh milk and meat distribution. Literature [11] proposed a hybrid ant colony algorithm combined with a local search operator to solve this problem. Literature [12] studies the integration problem of perishable food production and distribution with the time window under a time-varying road network and designs a hybrid genetic algorithm to solve it. Literature [13] builds a multivehicle and multiobjective LRP (logistic resource planning) model with the time window for this problem and proposes a hybrid solution algorithm of multiobjective particle swarm optimization and multiobjective neighbourhood search. Literature [14] proposed the LRP model, considering the influencing factors of fresh food quality attenuation.

3. Problem Description and Modelling

The location, vehicle information, volume, distribution of customer points, demand, time window, and other information of several candidate distribution centers are known. On the premise of satisfying the volume restriction and vehicle load limitation of distribution center, the site selection is made from the candidate distribution center and service customers are determined. Vehicles depart from selected distribution centers and each customer is served only once, minimizing the total cost while meeting customer needs and time window constraints.

This study establishes an optimization model to minimize total cost and maximize customer satisfaction:

$$\min z = \mathrm{BC} + \mathrm{TC} + \mathrm{PC}. \tag{1}$$

$$\max \frac{1}{t} \sum_{x \in C} \mu_x (n_x) \quad . \tag{2}$$

$$\begin{aligned}
\mathrm{BC} &= \frac{1}{\varpi} \sum_{b \in B} X_b K_b, \\
\mathrm{TC} &= \rho \sum_{x \in c} \sum_{y \in c} \sum_{z \in Z} W_z d_{xy} i_{xy}^z + \rho \sum_{x \in c} \sum_{y \in c} \sum_{z \in Z} F_z i_{xy}^z, \\
\mathrm{PC} &= \rho \sum_{x \in c} \sum_{z \in Z} \alpha_1 D_x \max\{\mathrm{ET}_x - n_x^z, 0\} \\
&\quad + \rho \sum_{x \in c} \sum_{z \in Z} \alpha_1 D_x \max\{n_x^z - \mathrm{LT}_x, 0\}.
\end{aligned} \tag{3}$$

TC consists of transportation costs and vehicle start-up costs. PC is the penalty cost of the delivery vehicle's advance or delay in meeting the customer's time window requirements.

Constraints are as follows:

$$\sum_{x \in C} D_x j_x^z \leqslant V^z, \forall z \in Z. \tag{4}$$

$$\sum_{x \in C} \sum_{y \in C} i_{xy}^z = |s_z| - 1, \forall z \in Z. \tag{5}$$

$$\sum_{x \in C_B} i_{xy}^z - \sum_{x \in C_B} i_{xy}^z = 0, \forall z \in Z, \forall y \in C. \tag{6}$$

$$\sum_{x \in C} \sum_{y \in C} i_{xy}^z = \sum_{y \in C} \sum_{x \in C} i_{xy}^z = 1, \forall z \in Z. \tag{7}$$

$$\sum_{x \in C_B} D_x K_b \leqslant O, \forall b \in B. \tag{8}$$

$$\mu_x (n_x) = \begin{cases} 0, & n_x \leqslant E_{x,} \\ \dfrac{n_x - E_x}{\mathrm{ET}_x - E_x}, & E_x < n_x \leqslant \mathrm{ET}_{x,} \\ 1, & \mathrm{ET}_x < n_x \leqslant \mathrm{LT}_{x,} \\ \dfrac{L_x - n_x}{L_x - \mathrm{LT}_x}, & \mathrm{LT}_x < n_x \leqslant L_{x,} \\ 0, & L_x < n_{x.} \end{cases} \tag{9}$$

Equation (1) represents the goal of minimizing the total cost, which consists of the distribution center depreciation cost, distribution cost, vehicle start-up cost, and the penalty cost for violating the time window. Equation (2) represents the goal of maximizing average customer satisfaction. Equation (4) represents the vehicle load constraint. Equation (5) indicates that each customer is served only by one vehicle. Equation (6) represents the elimination of subloops. Equation (7) indicates that the vehicle must leave after serving the customer. Equation (8) represents the volume constraint of distribution center. Equation (9) represents customer satisfaction membership function.

The above equations and coincidence are described below. C represents the customer set $\{xx = 1, 2, \ldots, t\}$ for delivery. B indicates distribution center set $\{bb = 1, 2, \ldots, w\}.C_B$ indicates the collection of distribution centers and customer points, that is, $C_B = C \cup B$. ϖ indicates the service life of the alternative distribution center. ρ represents the number of customer services per year. O indicates the volume of the distribution center. Z represents vehicle set $\{z|z = 1, 2, \ldots, z\}$. X_b represents the fixed construction cost of the distribution center b. F_z represents the start-up cost of vehicle z. W_z is the unit transportation cost of vehicle z. d_{xy} represents the distance between client x and client y. V^z represents the load constraint of vehicle z. D_x is the quantity demanded by customer x. n_{xy}^z indicates the travel time of the vehicle from customer x to customer y. n_x^z represents the time when vehicle z arrives at customer point i. S_z represents the collection of customer points served by vehicle z. $|S_z|$ represents the number of customer points served by vehicle z. ET_x indicates the earliest service start time expected by the customer. LT_x indicates the customer's expected start time of service. α_1 represents the penalty coefficient per unit time for the vehicle to service customer x before LT_x. α_2 represents the penalty coefficient per unit time of the vehicle's service to customer x after LT_x. E_x represents the earliest service start time that customer x can tolerate. L_x indicates the latest service start time that customer x can tolerate. δ is the confidence level of satisfying demands, that is, if the probability level of satisfying all demands reaches δ.

4. The Proposed Algorithm

In this study, the multidistribution network location and distribution path planning are studied. In order to reduce the difficulty of solving the problem, k-means clustering algorithm considering customer time and space is adopted. The multicenter path problem is transformed into a single center path problem. First, the k-means clustering algorithm is used to select the location of logistics outlets. The time-space two-factor algorithm designed in this study considering customers is used to determine the service customers of the distribution center. Finally, particle swarm optimization is used to optimize the path. The specific algorithm steps and flow are shown in Figure 1.

4.1. Algorithm Process Description

Step 1. Form an initial allocation. Use k-means clustering algorithm to determine the location of distribution center. Each customer point is divided to the nearest distribution center in turn to form the initial cluster.

Step 2. Calculate the time-space connectivity between clients. Suppose the time window of customer point x is $[ET_x, LT_y]$, the time window of customer point y is $[ET_x, LT_y]$. The service duration at customer point x is f_x. If the travel time from customer point x to customer point y is n_{xy}, the time-space linkage index χ_{xy} between customers is as follows:

$$\chi_{xy} = \begin{cases} 1, & ET_x + f_x + n_{xy} \leq LT_y; \\ 0, & \text{otherwise};. \end{cases} \tag{10}$$

If $\chi_{xy} = 1$ is displayed, customer x connects with customer y.

Step 3. Calculate and compare the time-space distance (space-time distance) index γ_{xb} of customer a and any distribution center b:

$$\gamma_{ab} = \frac{\sum_{n,y\in C} \chi_{xy}}{T_b D_{ab}}, \quad b \in B, \tag{11}$$

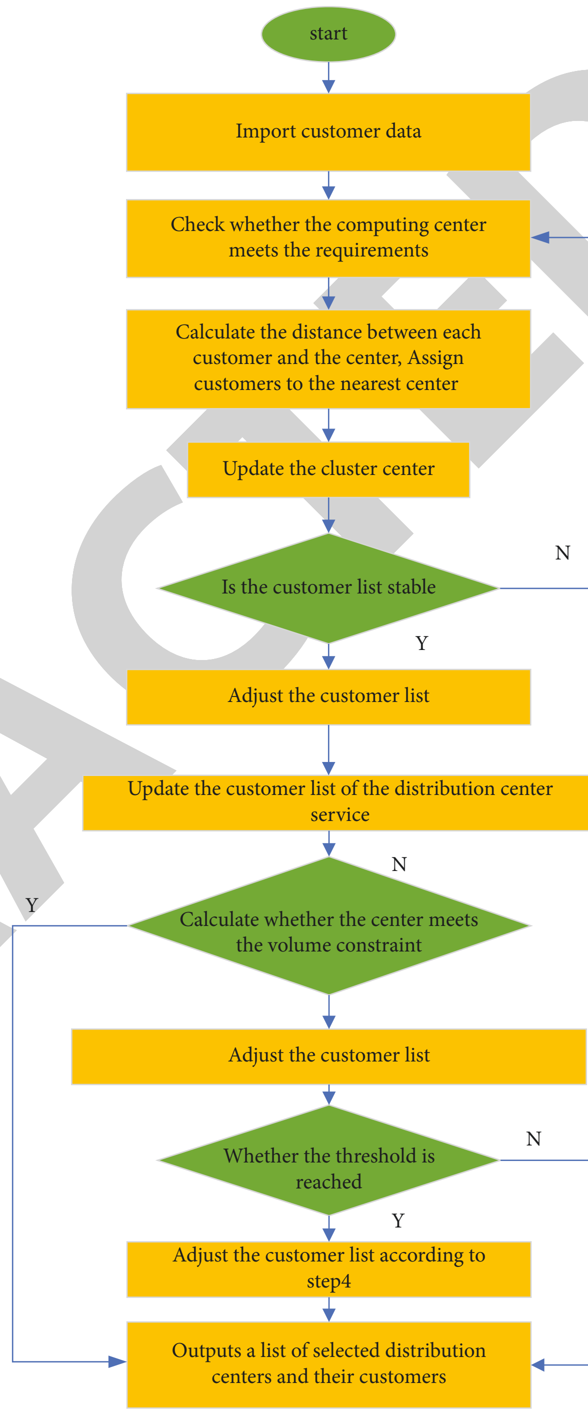


Figure 1: The specific algorithm steps and flow.

where T_b is the number of customers in distribution center b. D_{ab} is the distance between customer a and distribution center b. Compared with γ_{ab}, the customer is served by the largest distribution center in γ_{ab}.

Step 4. Check the capacity constraints of the distribution center and adjust the customer list for service in the distribution center:

$$J_b = \begin{cases} 1, & \dfrac{\sum_{x\in C_B} D_x K_b}{O} \leq 1, b \in B, \\ 0, & \text{otherwise}. \end{cases} \tag{12}$$

If $J_b = 0$, the sum of customer demands in the distribution center exceeds the upper limit of the distribution center capacity. Calculate the ratio of the spatiotemporal distance index of customer x to the sub-small spatiotemporal distance index $\gamma_{xb'}$ of the distribution center:

$$\varphi_x = \frac{\gamma_{xb}}{\gamma_{xb'}}. \tag{13}$$

It is assumed that the upper limit of the volume of the three distribution centers is 80, as shown in the Figure 2. At present, the total customer demands of the three distribution centers are 77, 75, and 86, respectively, and distribution center 3 exceeds the upper volume limit. According to step 4.4 rules of the design algorithm in this study, the customer 10 with the minimum spatiotemporal distance index and the demand greater than or equal to 6 in distribution center 3 is divided into the secondary distribution center 1. Then, the customer 3 in distribution center 1 whose demand is greater than or equal to 3 and whose spatiotemporal distance index is the smallest is divided into distribution center 2. The customer list after adjustment is shown in Figure 2. All distribution centers meet the upper volume limit.

4.2. Particle Swarm Optimization Fusion Algorithm

4.2.1. Standard Particle Swarm Optimization Algorithm. Particle swarm optimization is a swarm intelligence algorithm with a simple principle and fast search speed. The optimization idea of solving the optimal value is to simulate the foraging process of birds. Assuming that the initial values of velocity and position are randomly assigned in the solution space, W particles with space dimension D are searched. The idea of particle swarm optimization algorithm is to modify the position and velocity of particles through individual extreme pbest and global extreme gbest. It makes the particles move closer to the optimal solution. If the number of iterations is z, the update of particle velocity Q and position I is as follows:

$$Q^{z+1} = mQ^z + c_1 r_1 \left(U_{xd}^z - I^z\right) + c_2 r_2 \left(U_{ad}^z - I^z\right), \tag{14}$$

$$I^{z+1} = I^Z + Q^{Z+1}, \tag{15}$$

where m is the inertia weight. r_1 and r_2 are random numbers distributed in the interval [0, 1]. U_{xd}^z is the individual extreme value. U_{ad}^z is the global extreme value. c_1 and c_2 are usually taken as 2. When the population optimal solution reaches the preset range or Z is equal to the maximum number of iterations, the search is terminated.

4.2.2. Chaotic Particle Swarm Optimization Algorithm. Modern nonlinear theory interprets chaos as an aperiodic and irregular motion in a certain system. The logistic equation under the wormhole model is a typical chaotic system, which can be simplified as follows:

$$i_{t+1} = p i_t \left(1 - i_t\right). \tag{16}$$

When p takes 4 and i_t is a random number between 0 and 1, the output of the equation can traverse between 0 and 1 without repetition and quasi random. Therefore, by combining chaotic search with PSO algorithm, we can solve the problem that PSO algorithm is easy to fall into local optimization due to the strong randomness of particle initialization and evolution. Among them, the optimization of PSO using chaos can be divided into the following two points: one is to optimize the initial position and initial velocity using a chaotic sequence to improve the ergodicity and diversity of the population. Second, the optimal solution of the current population is searched by chaos, and the optimal result is used to replace the position of any particle in the current population. It can not only improve the convergence speed but also avoid the defect that it is easy to fall into local optimization.

4.2.3. Particle Swarm Optimization Algorithm. The chicken swarm algorithm (CSO) is a new bionic optimization algorithm, which mainly simulates the hierarchy and foraging behavior of chickens. The idea of CSO is to group chickens according to their types. Each Rooster can lead several hens and chicks into a group, and the hens in the group will forage under the guidance of the rooster. The chicks in the group can only forage around the corresponding hens, and information exchange is allowed between different groups. Roosters, hens, and chicks are classified from good to bad according to their fitness. In each round of search, the roosters, hens, and chicks in the group will be reselected.

The method of particle swarm optimization using the multi swarm nature of CSO is as follows:

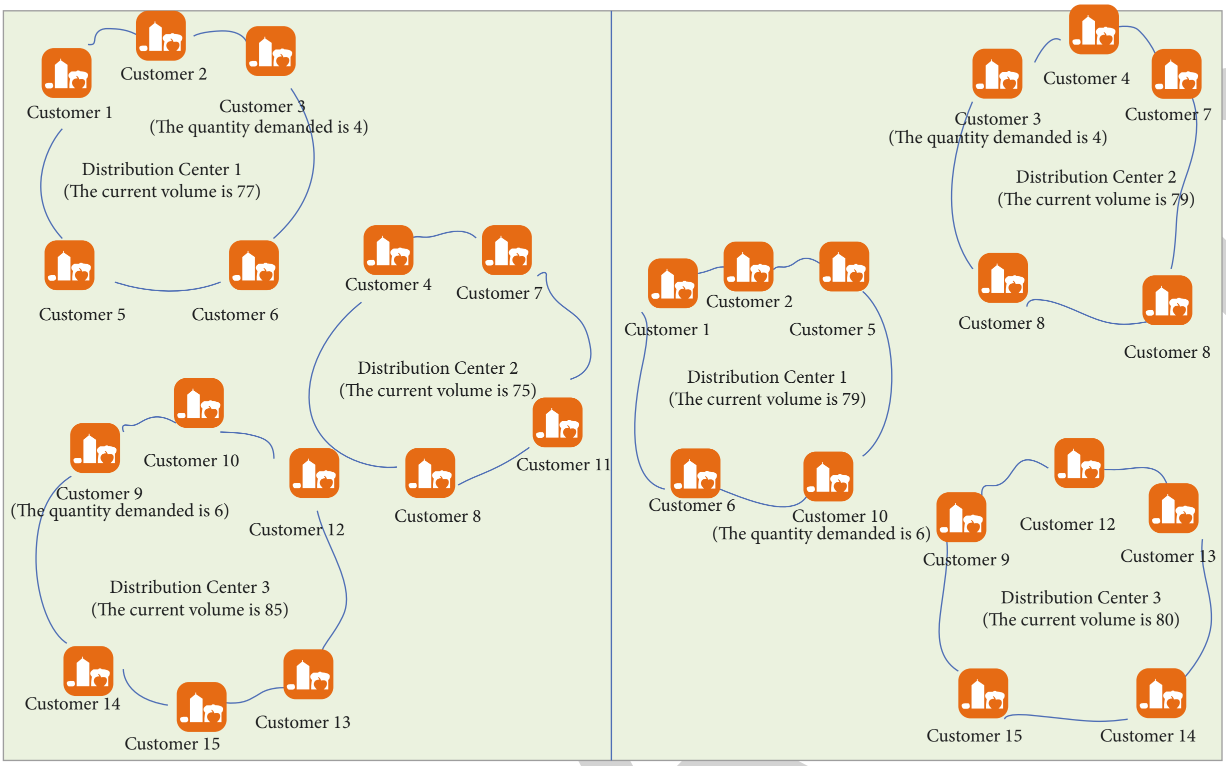
(1) In the first round of search, the fitness values of the population are sorted from small to large, and then all particles in the population are divided into three categories: G particle (rooster), H particle (hen), and C particle (chick). Groups are grouped according to rules, and the other rounds update the member types in the group by comparing the fitness values in the group, without changing the order of group numbers.

(2) When updating the speed and position of particles in each round, class G particles are excellent individuals in the group, and the updating formula is the same as formulas (14) and (15).

Class H particles search under the guidance of class G particles and absorb the experience of other groups. The iterative formula of its speed and position can be changed to

$$\begin{aligned} Q_H^{z+1} &= mQ_H^z + c_1 r_1 \left(U_{xd}^z - I_H^z\right) + c_2 r_2 \left(U_{ad}^z - I_H^z\right) \\ &\quad + c_3 r_3 \left(I_f^z - I_H^z\right), \\ I_H^{z+1} &= I_H^z + Q_H^{z+1}. \end{aligned} \tag{17}$$

Among them, $U_{a\,d}^z$ changes from the global optimization of equation (1) to the intragroup optimization, r_3 is a



(a) (b)

FIGURE 2: The adjusted customer list. (a) Before customer list adjustment. (b) After customer list adjustment.

random number of 0–1, I_f^z is the optimal position of other groups, and c_3 is usually taken as 2. Class C particles can only be searched near class H particles, and their position iteration formula is as follows:

$$I_C^{z+1} = I_C^z + F_{\mathrm{FL}} \times \left(I_{Ha}^z - I_C^z\right), \tag{18}$$

where I_{Ha}^z is the position of H particle corresponding to C particle and FL is usually taken as 0.5.

5. Experimental Results and Analysis

Taking the actual problem of location and route of fresh electricity supplier in a city as a case study. The number of customers and their locations of the e-commerce service are relatively stable. The current distribution center capacity is limited and cannot meet the growing customer demand. In order to reduce distribution costs and improve customer satisfaction, the e-commerce company plans to open new distribution outlets and hire professional distribution teams to make distribution plans. The alternative distribution center of the e-commerce has been basically determined. The number and location of customers served by the company are relatively stable, keeping at about 60 customer points for a long time. The location, construction cost, and processing capacity of alternative distribution centers are listed in Table 1. A one-day order, a total of 60 customers (numbered 1, 2, ..., 60), and rent 8 operation vehicles with the same specifications from the distribution company are selected. The maximum load of the vehicle is 100 kg, and the speed of the vehicle is 60 km/h. The fixed cost of the vehicle is 150 yuan/vehicle, the transportation cost per unit distance is 0.57 yuan/km, and the value loss coefficient of the unit product per unit time is 0.5 yuan/hour. The service time of the vehicle at the customer point is related to the order quantity of the customer, and the service time of the goods per unit weight is 0.02 hours. The opportunity cost loss coefficient due to arriving earlier than the customer's expected time window is 10 yuan/hour. The penalty cost factor for arriving late in the customer's expected time window is 20 yuan/hour. All algorithms were run using python3.8.2 on an ASUS W518

TABLE 1: Information table of alternative distribution centers.

Number	Longitude	Latitude	Yuan/day	Storage capacity (kg)
1	238.28336	50.286358	3670	530
2	238.21983	50.179897	6086	870
3	238.36284	50.220049	3966	640
4	238.29205	50.164098	6320	1110
5	238.28788	50.190539	4094	570
6	238.29852	50.148184	6051	870
7	238.24347	50.195803	3671	540
8	238.41584	50.189083	6620	1260
9	238.37028	50.165667	4380	670
10	238.38666	50.222344	6210	1010
11	238.36424	50.239137	3790	650
12	238.2029	50.202132	3300	660
13	238.27348	50.223855	4285	580
14	238.29489	50.255086	4050	550
15	238.27355	50.256708	3870	560

computer with an Inter(R) Core(TM) CPU i7-6700 K@ 4 GHz, 8 GB RAN 64 bit, Windows 10 operating system.

The location, demand, and time window information of customers are listed in Table 2. The customer time window is a fuzzy variable represented by a trapezoidal fuzzy number. Due to the limited layout, Table 3 only lists the information related to the top 20 customers.

5.1. Taguchi Analysis. Algorithm parameters have a great influence on algorithm performance, and good parameter settings can greatly improve algorithm performance. Therefore, it is very important to determine reasonable parameter values. Taguchi design is an excellent parameter design method, originating from the field of quality management, which aims to determine the optimal level combination of parameters with the minimum number of tests [15]. The algorithm has been widely used for its superior reliability, reproducibility, and simple analysis. Taguchi method mainly measures parameters by the signal-to-noise ratio (S/N). The larger the signal-to-noise ratio is, the more reasonable the corresponding parameter settings are. In addition, Taguchi's target values were divided into three groups: "smaller is better", "bigger is better," and "nominally best". Since the objective function of this study is to minimize the total cost, the type of "the smaller the better" is selected, and the corresponding SNR calculation equation is as follows:

$$\frac{S}{N} = -10 * \log_{10}\left(\sum_{n=1}^{t}\frac{F_n^2}{t}\right), \tag{19}$$

where n is the times of algorithm execution at each parameter level. F_n is the response value, that is, the objective function value of experiment n.

Parameter settings of PSO algorithm mainly include iteration times, population size, learning factor, inertia weight, crossover probability, and mutation probability. The nonlinear dynamic inertia weight coefficients, $m_{\max} = 0.9$ and $m_{\min} = 0.4$, are used in this study. According to previous experience, the learning factor (C) was set as 2, and the number of generation (Gen) was set as 500. Population size (POP), crossover probability (CP), and mutation probability (MP) were optimized by Taguchi analysis to determine the optimal values. Considering the optimization of orthogonal experimental design, three levels were taken for each parameter: the population size was 30, 40, and 50, respectively. The crossover rates were 0.75, 0.80, and 0.85, respectively. The variation probability was 0.10, 0.15, and 0.2, respectively.

Taguchi analysis was carried out using an A-type orthogonal test table, and each horizontal combination was repeated for 5 times. Table 3 lists the orthogonal test results of Taguchi analysis, the experimental value of total cost under each scheme, and corresponding signal-to-noise ratio and mean value.

5.2. Convergence Analysis of the Algorithm. In this study, the convergence of PSO algorithm is analyzed and verified by using two test functions combined with the test results of benchmark cases. For small and medium scale cases, the algorithm converges to the global optimal solution. There are 11 calculation examples in 24 Prins case sets [16] and 19 Barreto case sets [17], respectively. The gap between the solution results of PSO and BKR is 0%, indicating that the algorithm has converged the global optimal solution at this time. When solving large scale, the average difference between the optimal solution of the algorithm and BKR is only 1.16% in the Prins case set. In Barreto case set, the average difference between the optimal solution of PSO and BKR is 2.07%, excluding cases per 183–318 × 4 and per 183–318 × 4. The average difference between the optimal solution of all large-scale cases in the two case sets and BKR is 6.02%. Therefore, the PSO algorithm can converge to obtain an approximate optimal solution when solving large-scale problems. Finally, two common test functions are used to test as follows:

$$
\begin{aligned}
T_1 &= \begin{cases} \min F(i,j) = (i_1 - 30)^2 + (i_2 - 20)^2 - 20 j_1 + 20 j_2, \\ \text{s.t.} - i_1 - 2i_2 + 30 \le 0, \\ i_1 + i_2 - 25 \le 0, \\ i_2 \le 15, \\ \min f(i,j) = (i_1 - j_1)^2 + (i_2 - j_2)^2, \\ \text{s.t. } 0 \le j_1 \le 10, \\ 0 \le j_2 \le 10, \end{cases}, \\
T_2 &= \begin{cases} \min F(i,j) = -i_1^2 - 3i_2 - 4j_1 + j_2^2, \\ \text{s.t. } (i_1)^2 + 2i_2 \le 4, \\ i_1 \ge 0, i_2 \ge 0, \\ \min f(i,j) = 2i_1^2 + j_1^2 - 5j_2, \\ \text{s.t. } i_1^2 - 2i_1 + i_2^2 - 2j_1 + j_2 \ge -3, \\ i_2 + 3j_1 - 4j_2 \ge 4, \\ j_1 \ge 0, j_2 \ge 0. \end{cases}
\end{aligned} \tag{20}
$$

For the test function, T_1 obtains its optimal solution $(i_1, i_2, j_1, j_2) = (20, 5, 10, 5)$ through the PSO algorithm. $F = 225$ is the upper target value. $f = 100$ is the lower target value, and the convergence curve is shown in Figure 3. The optimal solution of T_2 is $(i_1, i_2, j_1, j_2) = (0.4991, 1.8733, 4.1828, 2.6054)$. $F = -15.8120$ is the upper target value, $f = 4.9668$ is the lower target value, and the convergence curve is shown in Figure 4. The convergence curve shows the value of the objective function under different iterations. For both test functions, the optimal solution is obtained in the 6th or 7th generation, which further verifies the convergence and high efficiency of PSO algorithm.

5.3. Sensitivity Analysis of the Algorithm. In the case solution of this study, the confidence level δ is assumed to be 0.9. In order to get a further conclusion, we calculate each cost at each confidence level δ from 0.05 to 0.95 successively. The algorithm in this study is used to calculate the optimal solution that meets the respective demand conditions at

Table 2: Customer information table.

Number	Longitude	Latitude	Demand (kg)	Time window			
				E_x	ET_x	LT_x	L_x
1	238.254288	50.254953	9	0	0	4	6.5
2	238.225254	50.242306	11	0	1	4	6.5
3	238.267942	50.20603	8	0	1	4	8
4	238.254144	50.244656	10	0	0	4.5	6
5	238.281165	50.217452	10	0	1	4	5.5
6	238.275703	50.210734	10	0	1	4	5
7	238.247389	50.213757	11	0	0	3	6
8	238.212031	50.221371	12	0	0	5	7
9	238.319684	50.22417	11	0	0	4	6.5
10	238.334057	50.230328	12	0	0.5	5	6.5
11	238.300568	50.233924	1I	0	0.5	5.5	6.5
12	238.346561	50.198303	12	0	1	5.5	6
13	238.295681	50.251595	9	0	0	6	6.5
14	238.315516	50.262115	12	0	1	4.5	6
15	238.247245	50.258086	12	0	1	4.75	5.5
16	238.326008	50.248126	10	0	1	5	5.5
17	238.316163	50.231056	10	0	1	5.5	6.5
18	238.304449	50.216501	8	0	0.5	4	6
19	238.337866	50.207654	9	0	1	6	7
20	238.30912	50.192198	8	0	1	4.5	6

Table 3: The experimental design analysis.

No.	POP	CP	MP	Run1	Run2	Run3	Run4	Run5	SNR	Mean
1	30	0.75	0.1	9597.56	9590.28	9465.21	9533.2	9669.39	−78.5	9571.128
2	30	0.8	0.15	9556.5	9558.07	19535.14	13524.6	9478.69	−81.1501	12330.6
3	30	0.85	0.2	19520.59	9498.59	9469.82	19438.21	9542.49	84.0789	13493.94
4	40	0.75	0.15	9580.7	9547.59	19555.63	9527.94	13616.2	−81.1735	12365.61
5	40	0.8	0.2	9557.88	9628.05	9652.27	15679.08	9592.97	81.8166	10822.05
6	40	0.85	0.1	9848.5	5930.95	15930.62	9560.2	12706.48	−79.9647	10795.35
7	50	0.75	0.2	5369.27	9619.34	5879.3	9767.54	9527.24	−77.22	8032.538
8	50	0.8	0.1	1557553	9551.91	1891155	13516.46	5777.61	−81.5453	695510.8
9	50	0.85	0.15	6000.8	5833.96	9608.31	9525.3	9589.91	−78.5	8111.656

their respective confidence levels and observe their changes, and the change curve as shown in Figure 5 can be obtained.

Through the analysis, the change trend of total cost and logistics cost is basically the same, and logistics cost accounts for a relatively high proportion. Therefore, it is very important for enterprises to choose where to establish distribution centers and establish several distribution centers from the perspective of overall cost when making location decisions.

In addition, it can be seen from Figure 5 that in the process of increasing confidence level, the input cost of enterprises is also gradually increasing. The confidence level essentially represents the percentage of shipments that can meet demand at the retail end. When the demand is uncertain, enterprises should pay more and more costs to meet the demand of the retail end. Meanwhile, the impact on cost can be further analyzed by adjusting the standard deviation σ. When the standard of quantity demanded changes, the actual quantity demanded changes. Standard deviation is often used to evaluate the degree of shadow of uncertain factors, and the greater the standard deviation is, the greater the uncertainty is.

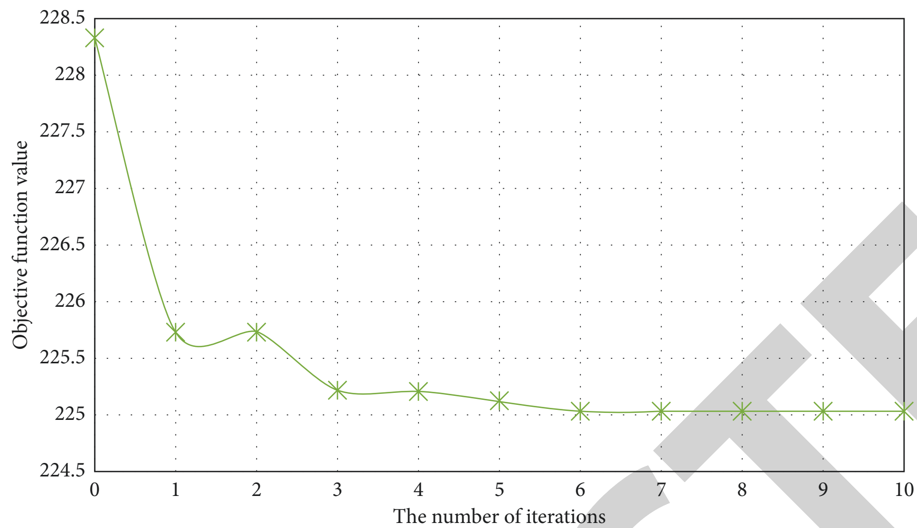


Figure 3: Convergence curve of test function T1.

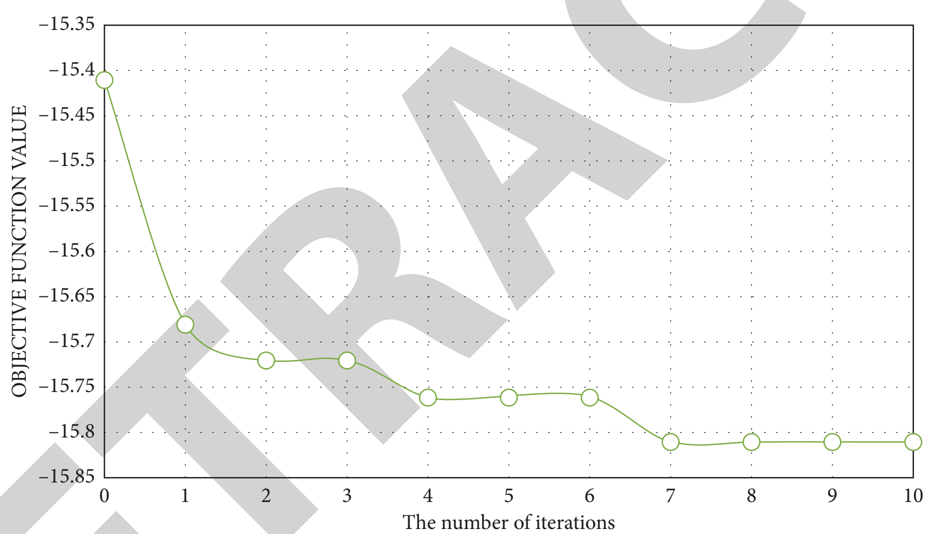


Figure 4: Convergence curve of test function T2.

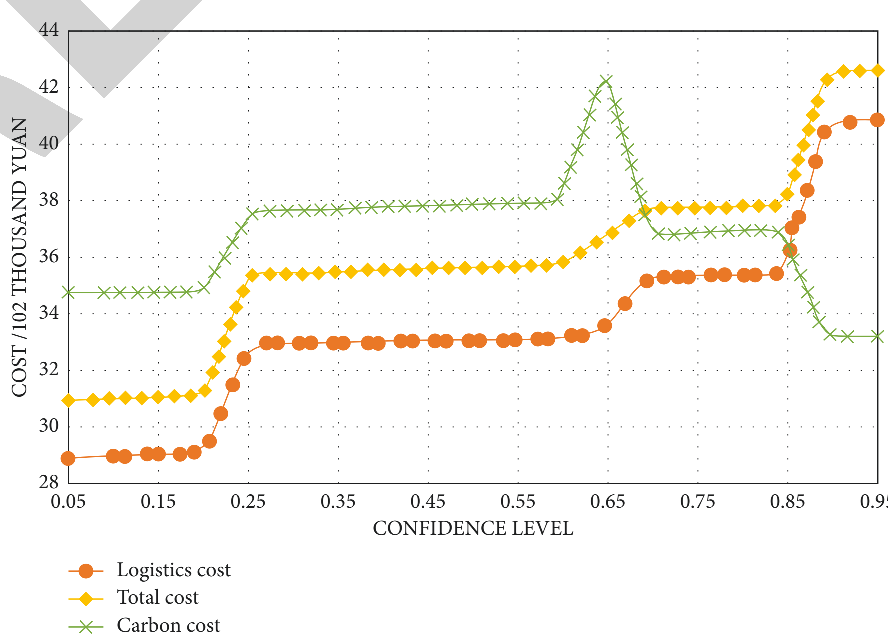


Figure 5: Cost change analysis.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9843154, 1 page
https://doi.org/10.1155/2023/9843154

Retraction

Retracted: Optimization Model of Logistics Task Allocation Based on Genetic Algorithm

Security and Communication Networks

Received 25 July 2023; Accepted 25 July 2023; Published 26 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Wang and J. Gao, "Optimization Model of Logistics Task Allocation Based on Genetic Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5950876, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5950876, 13 pages
https://doi.org/10.1155/2022/5950876

Research Article

Optimization Model of Logistics Task Allocation Based on Genetic Algorithm

Xueli Wang and Jingjuan Gao

Shijiazhuang Institute of Railway Technology, Shijiazhuang 050041, China

Correspondence should be addressed to Jingjuan Gao; gaojing0635juan@163.com

Received 15 March 2022; Revised 6 April 2022; Accepted 19 April 2022; Published 2 May 2022

Academic Editor: Chin-Ling Chen

In order to improve the efficiency of logistics task allocation, the rationality and algorithm of the logistics cloud task scheduling model based on genetic algorithm are proposed in this paper. Firstly, the basic principle of genetic algorithm is introduced, the logistics cooperative distribution model is constructed, and the judgment mathematical model of the transfer point of the logistics distribution demand point is constructed. Genetic algorithm is used to solve the logistics distribution path planning model, and the model is simplified. The complex multiobjective optimization problem is transformed into a single-objective optimization problem through preference vector. The genetic algorithm and open-source algorithm on Python are used to simulate the model proposed in this paper. From the change curve of the objective function, after 100 generations of iteration, the value of objective function increases rapidly from 30 to 130 and slowly from generation 5 to generations 40 to 130. Subsequently, the 40th generation to 60th generation were rapidly upgraded to 160. Finally, the 60th to 100th generations are basically stable at about 170. The cost in the scheduling process decreases gradually with the increase of the number of iterations of the algorithm, from the initial unit cost of nearly 200 to 120. Then it gradually decreases to about 80. Genetic algorithm shows the ability of efficient and accurate solution in this 100-generation iteration. The genetic algorithm is used to solve the problem. The algorithm parameters are as follows: population size pop size = 300, maximum number of iterations max gen = 200, crossover probability PC = 0.8, and mutation probability PM = 0.1. Using the data in this paper and substituting it into the model established in this paper, the following distribution scheme is obtained: p the minimum distribution cost is 601.58 yuan, the distribution vehicle is 5, and the total mileage is 477.41. After using the algorithm to optimize the path, the path interleaving is greatly reduced, and the vehicles do not take the repeated route, which can greatly save the cost. After calculation, the total mileage after optimization is 74.8% lower than that before optimization, and the cost is significantly reduced by 72.8%. To sum up, the last kilometer distribution algorithm proposed in this paper can greatly reduce the cost of logistics resource scheduling, which has obvious research significance.

1. Introduction

At present, with the rapid advancement of intelligent manufacturing in the automotive industry, especially under the background of China's vigorous promotion of the Internet of things, big data, cloud computing, and intelligent equipment, the use of automatic logistics system in the process of intelligent manufacturing has become an important measure for the company to improve its comprehensive strength and plays an important role in the intelligent construction of production management process. How to effectively strengthen the application effect of automated logistics system has become one of the key issues of enterprise intelligent manufacturing [1]. Logistics plays a more and more important role in automobile production. Logistics cost is an important part of product cost, and it also affects product quality to a great extent. Therefore, the level of logistics planning and implementation directly affects the competitiveness of enterprises [2]. The requirements for logistics are higher in the whole vehicle production plant. Because the automobile is a highly integrated product, nearly 10000 parts of the whole vehicle are basically distributed to the final assembly line of the factory through logistics except the body of the vehicle [3]. With the increasingly fierce competition in the automotive industry and the increasing pressure on quality, cost, and efficiency, the logistics of the final assembly workshop has developed rapidly from rough logistics to lean logistics. In order to effectively improve the independent matching ability of the overall supply and demand data and realize the development and construction of the future logistics system, the informatization and networking of the logistics system have developed significantly. The continuous production of relevant automation software and website platform has effectively alleviated the conflict between supply and demand of contemporary logistics information, but there are still many defects: first, the sharing rate of logistics resources is very low, and many logistics resources have not been effectively used, resulting in the low utilization rate of logistics resources; the second is the lack of basic flexibility of logistics services (such as failure of service links and lack of replacement efficiency); third, users have relatively little access to logistics information, which makes it difficult for them to explore the most ideal logistics services; fourth, the current logistics information lacks basic transparency and fails to solve the personalized needs of users. If these problems continue to exist, they will directly or indirectly lead to the rise of logistics costs and the significant reduction of service quality [4]. Therefore, it is necessary to optimize the logistics allocation, as shown in Figure 1, the transportation management system of multitask scheduling. Based on the current research, this paper uses genetic algorithm to solve the logistics distribution path planning model, simplifies the model, and transforms the complex multiobjective optimization problem into a single-objective optimization problem through preference vector. Fitness function value of genetic algorithm. The genetic algorithm and open-source algorithm on Python are used to simulate the model proposed in this paper, referring to the model established in this paper, combined with the relevant genetic algorithm to realize the effective design and solution, and through the python software to realize the simulation, so as to verify the feasibility and effectiveness of the model and algorithm established in this paper.

2. Literature Review

Aiming at this research problem, Lammie and others studied the influence of the redistribution of picking bits on the total picking time of the array automatic picking system in the serial merge picking mode [5]. Bu and others studied a location redistribution strategy that optimizes the location and picking path in two stages, constructed the total picking time model of stacker, studied the influence relationship between the path and the total picking time model, and solved it by heuristic algorithm [6]. Zhong and others proposed that order picking has periodicity. By studying the order characteristics of the distribution center, the order trend is mined and predicted, and based on this, the picking position allocation of goods is optimized [7]. Cao and others proposed comparing the space required to store a certain type of goods with the average outbound volume of the goods as an indicator of picking bit allocation [8]. Li and others analyzed that when manual picking and automatic picking systems are used for partition picking and parallel picking, the manual picking area or unallocated area will affect the picking path of operators and then affect the picking efficiency [9]. Martin and others studied the influence of location allocation on picking efficiency in this type of automatic three-dimensional warehouse with multiple lanes, selected several influencing parameters, i.e., cargo correlation and shipment volume, established the picking time model, and selected heuristic algorithm genetic algorithm to search the optimal solution [10]. Soleimanpour Moghadam and others compared the cost problem of manual picking and automatic picking and took a pharmaceutical distribution center as an example to find the optimal solution by using greedy algorithm [11]. Gao and others simulated the static and dynamic allocation model with the goal of minimizing the waiting time of ships and solved the model by Lagrange relaxation method [12]. Zhou and others used agent technology to simulate the ship transportation strategy of container port and compared and evaluated different strategies [13]. Reis and others transformed the NP berth allocation problem into a multistage decision-making program and proposed the improved directional search plan, two-stage node quality evaluation, and random node selection criteria by using the random directional search algorithm [14]. Based on the current research, this paper uses genetic algorithm to solve the logistics distribution path planning model, simplifies the model, and transforms the complex multiobjective optimization problem into a single-objective optimization problem through preference vector. Fitness function value of genetic algorithm. The genetic algorithm and open-source algorithm on Python are used to simulate the model proposed in this paper, referring to the model established in this paper, combined with the relevant genetic algorithm to realize the effective design and solution, and through the python software to realize the simulation, so as to verify the feasibility and effectiveness of the model and algorithm established in this paper.

3. Method

3.1. Basic Principle of Genetic Algorithm

3.1.1. Generation of Genetic Algorithm. Genetic algorithm (GA) is a randomized search method that simulates natural evolution. GA is the survival process of the fittest that represents the problem as a "chromosome." Through the generation by generation evolution of the "chromosome" group, through operations such as replication, selection, crossover, and mutation, it finally converges to the individual "most suitable for the environment," so as to obtain the optimal solution or approximate optimal solution of the problem [15, 16]. Because GA directly takes the objective function as the search information and does not need high value information such as gradient, at the same time, it adopts the adaptive random search technology (imitating the "survival of the fittest" law in the biological world) and uses the search information of multiple search points, so it has high flexibility, parallelism, and strong universality.

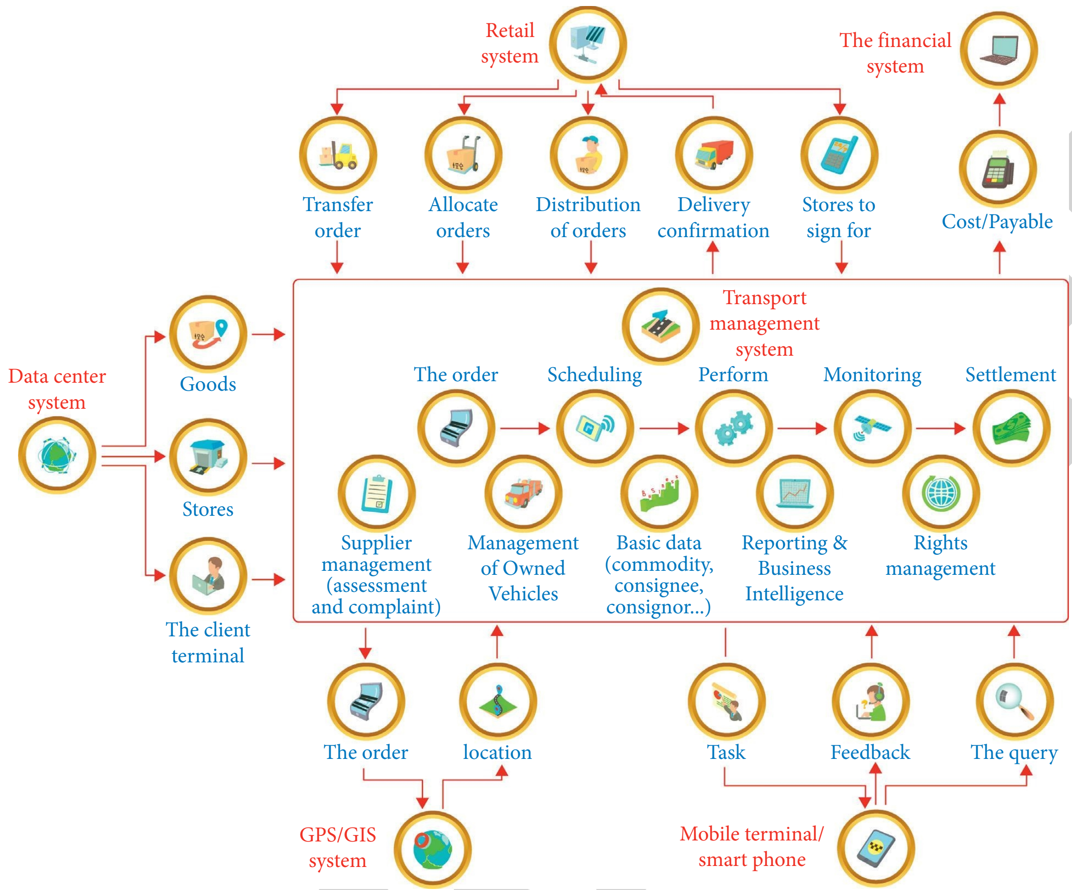


Figure 1: Transportation management system of multitask scheduling.

3.1.2. Advantages of Genetic Algorithm. Compared with precise algorithms and relevant heuristic algorithms, relevant genetic algorithms have strong robustness when dealing with problems in a certain optimization space. The application program is relatively simple and convenient, and the scope of use is wider. Then, they can establish efficient interception with big data and carry out optimization and processing by integrating the functions of cloud computing. Its corresponding advantages are as follows: it has the corresponding self-organization and management ability and is outstanding in adaptability and intelligence. In dealing with related optimization problems, the first step of genetic algorithm needs to create corresponding coding scheme, matching function, crossover, and other parameters [17]. Then, carry out the links of gene variation and combination according to the fitness of different individuals and the similar conditions of organizational framework, so as to obtain another population. The objective function corresponding to the problem to be handled acts as the corresponding search task. For the previous algorithms, when dealing with and calculating problems, in addition to mastering the value of the objective function, they must also have a series of auxiliary information. It is generally required to expand the operation and derivation around the objective function [18]. However, in the new genetic algorithm, the target to be searched can be directly divided into the corresponding operation space with more ideal fitness value. On this basis, it can comprehensively improve the search ability of the overall search of the algorithm, speed up the operation efficiency of the algorithm, and compress the time required in the search link as much as possible. The operation object is directly related to the decision variables that complete the coding link. In the previous algorithms, for the problem processing in the combinatorial field, the solution and operation are usually carried out according to the real value corresponding to the decision variable of a specific object. However, in the new form of genetic algorithm, the corresponding coding task is carried out by using the coding mode matching with relevant variables. Under this arrangement, if there is no value in the follow-up, it can be processed efficiently. With hidden parallelism, it can be integrated with popular information technologies such as big data and cloud computing, which can greatly improve the efficiency of processing tasks. In previous search algorithms, the link of operation and solution generally starts from a single point, and the optimal answer is easy to be partial and one-sided. For genetic algorithm, it does not start from a single point, but search in a population of a certain size, and search in different directions at the same time [19]. The parallelism of this aspect is mainly reflected in the following: firstly, it can be fully integrated with big data and cloud computing, and then it has internal parallelism in computing information; in addition, searches are performed in different directions to satisfy a certain degree of parallelism, guiding the development of all search links according to their corresponding possibilities. In the previous algorithms, the search is generally carried out from a point part along a specific direction. In such a situation, it is likely that the best answer cannot be obtained. For heredity, it is not from a single point, but to search in a certain size of population, and search in different directions at the same time. According to the relevant fitness value and probability, the direction of search is clearer, which can avoid obtaining partial optimal solutions, and improve its global search performance to a great extent.

3.1.3. Basic Principle and Program of Genetic Algorithm. Genetic algorithm is significantly different from previous search algorithms. For the former, the search is usually carried out in a series of randomly obtained preliminary solutions, and a group of solutions is generally called group. Because different individuals in the population are the corresponding solutions of related problems, they are generally called chromosomes. This kind of chromosome will be continuously adjusted and optimized in the subsequent iterative links, which is called genetic processing. Through the steps of crossover and mutation, the subsequent chromosomes are obtained, that is, offspring. For the quality of chromosomes, it is usually considered with the aid of the index of fitness. According to the scale of fitness, a corresponding number of individuals are selected from the previous generation and future generations, which play the role of the next generation group and then continue to develop and evolve. When the long actual development and change is completed, the algorithm will get the chromosome with the most ideal convergence effect. At this time, the chromosome obtained is the optimal solution or the second optimal solution [20]. In the aspect of genetic algorithm, the value of fitness is used to analyze the specific situation of the optimal solution that different individuals can obtain after the optimization link. The function that measures the fitness of different individuals is the fitness function. The definition of this function is usually related to the problems that need to be solved. Table 1 is the concept often involved in genetic algorithm.

Genetic algorithm corresponds to four basic processing programs: coding, screening, crossover, and mutation.

Coding: how to express the feasible explanation of relevant problems in the algorithm, that is, to transform the feasible solution of a specific problem into the search field that the algorithm can calculate and operate in the corresponding space. This transformation method and strategy is coding. The coding mode directly affects the comprehensive performance of the algorithm, as well as the initialization setting steps and the design of different operators.

Selection: the key task of this part of work is to obtain the ideal individual from the current whole, so that it can continue to develop subsequent offspring as a parent generation. According to the fitness status of different individuals, in terms of population, select ideal individuals from the previous generation according to the corresponding principles and strategies, and then transfer them to the subsequent generation through genetic steps. The screening standard is to obtain individuals with stronger adaptability. In this way, we can make greater contributions in the subsequent reproduction and sending letters, and we are more likely to get more offspring. Such a procedure obviously embodies the concept of survival of the fittest, which is highly similar to the meaning of biology.

Crossover: this work is the most critical operation in genetic algorithm. With the help of cross-processing, individuals in subsequent generations can be obtained, and these individuals obviously inherit a series of attribute characteristics of the previous generation. Cross-processing is to complete the corresponding matching operation for different individuals in the group. For different individuals, some chromosomes they have are exchanged and shared according to the corresponding probability. This part embodies the concept of information exchange.

Mutation: mutation processing is to adjust some positions on different coding strings according to a very small possibility. For example, 0 in binary coding is adjusted to 1, so as to get a new individual. Although this kind of algorithm is only an auxiliary means to obtain new individuals, it is also a very key link, which is directly related to the search performance of genetic algorithm in local space. Crossover and mutation operators work together to complete the follow-up global search and local search. Like the biological field, the possibility of variation in GA is very low, and its corresponding value is generally only 0.001-0.01.

3.2. Construction of Logistics Cooperative Distribution Model

3.2.1. Problem Description. Before the provider of logistics resources makes specific distribution, it is necessary to refer to the actual tasks and distribution needs to carry out planning for the relevant tasks of distribution, and it is necessary to calculate the spare space of relevant vehicles in the distribution link and publish the basic type, capacity, and cargo information of acceptable goods through the platform at the low price exceeding the cost price. The remaining providers further develop relatively low-cost solutions by effectively capturing and reading the platform data and select enterprises with stronger capabilities with reference to relevant needs to achieve effective cooperation between the two. On the basis of relevant cooperation, enterprise 1 has evolved from "enterprise" to "logistics service demander," and enterprise 2 has further evolved from the initial "irrelevant enterprise" to "logistics resource provider." In such cases, logistics companies may have dual roles. On the one hand, they may be the basic demander, but also the service provider. Based on the logistics cooperative distribution problem, the first is to carry out research on the relevant distribution demand points, so as to identify whether the relevant transshipment operations should be carried out. The focus of the judgment is that, for distribution demand points, when the cost of their own distribution activities exceeds the cost of other enterprises, or their own ability fails to solve the distribution demand, they can choose the transshipment scheme. At the same time, after determining the demand point and carrying out the transfer

Table 1: Correspondence between the basic concepts of biological genetics and the role of genetic algorithm.

Meaning of biological heredity	Functions of genetic algorithm
Survival of the fittest	After the algorithm is completed, the answer of the best target value is very likely to remain
Group	A group of solutions selected (the number of solutions at this time represents the size of the population)
Mating	Those who obtain a new set of solutions with the help of the cross principle will not go
Population	A set of solutions obtained according to the value of fitness function
Variation	A program in which a single component corresponding to the code changes
Chromosome	Encoding of solution (string and other related contents)
Gene	Eigenvalue of single component in solution
Threshold	Set the criteria for completing the search
Adaptability	Fitness function data
Individual	Solution

meeting, carry out the corresponding path planning for the overall logistics distribution. See the identification in Figure 2 for the actual process.

The first step is to carry out effective judgment on relevant demand points at the distribution planning level. Before carrying out relevant distribution business, the "provider" shall carry out planning for the distribution task with reference to the actual task and distribution demand and disclose the types, capacity, and relevant information of goods that can be handled through the platform at a relatively low price. The rest of the "demanders" rely on the platform to view and read information, further develop companies with lower transportation costs, and select reliable enterprises for cooperation with reference to actual needs. The "provider" in this stage belongs to the relevant cooperative distribution company. Its basic distribution demand point location belongs to the determined mode: carry out research on the distribution points of the relevant "demander," and further explore the enterprises with relatively lower distribution cost to realize the corresponding cooperative distribution by comparing the cost differences between the "demander" and the "provider."

The second step is the route planning layer, so as to judge the most ideal distribution route. Referring to the judgment formed by the relevant distribution planning layer for the actual distribution point, this level will be started. According to the actual demand points, realize the supporting renumbering operation, carry out the basic path planning, and execute the relevant distribution operations at the same time. There are a large number of distribution companies in the platform. At the same time, each enterprise has I distribution tasks. Referring to the sharing and related predictability of the platform foundation, it is concluded from the relevant path planning for logistics distribution that the "provider" has significant fluctuations in the cost in a specific distribution area with the continuous increase of time and related logistics tasks. Relying on the comparison of task I cost, the "logistics service demander" further concludes that the distribution of the "provider" is helpful to control the overall cost, so the "provider" is responsible for the implementation of this logistics task. Then, the task at this stage belongs to the category of transshipment and

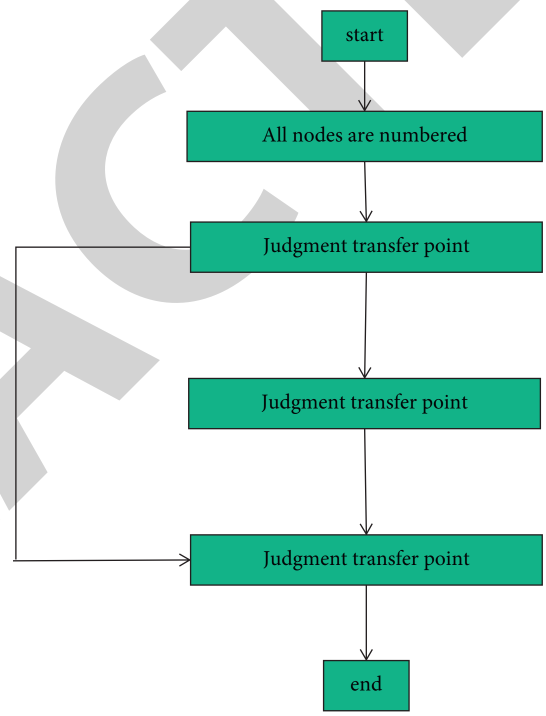


Figure 2: Flow chart of path planning algorithm.

distribution task. In the actual cooperative distribution system, the description adopted belongs to the specific location information of the known distribution center and logistics service demander, the basic data of relevant distribution centers and vehicles (load, limit distance, etc.), the demand information of the demander, and the unit cost of relevant nodes.

3.2.2. Model Assumptions. Referring to the basic characteristics of relevant distribution point problems, this paper obtains the following basic assumptions: the goods distributed are the affordable category of time window; relevant distribution enterprises have redundant distribution capacity and can carry out cooperative distribution of logistics tasks; it is assumed that in the relevant transshipment procedures the management expenses due to the platform are not included.

3.2.3. Variable Definition. Among them, Tables 2 and 3 show variable lists, and Table 4 shows the alphabetic list.

3.3. Model Establishment

(1) The mathematical model for judging the transfer point of logistics distribution demand point is as follows:

Objective function:

$$\max f = \sum_{p=1}^{G} c_{pj} d_{pj} - \sum_{q=1}^{G} \sum_{p=1}^{G} c_{pq} d_{pq} - \sum_{q=1}^{G} c_j^q j = (1, 2, \ldots, N). \tag{1}$$

Constraints:

$$\begin{aligned} c_{pq} &= 0, \\ p &= q, \end{aligned} \tag{2}$$

$$\sum_{p=1}^{G} y_{jp} = 1, \tag{3}$$

$$\sum_{j=1}^{M} q_j y_{jp} \leq Q, \tag{4}$$

$$y_{jp} = \begin{cases} 1, \text{The distribution of logistics task J is completed by company P}, \\ 0, \text{other}. \end{cases} \tag{5}$$

Equation (1) is the objective function. The first item represents the cost of the logistics company to distribute a logistics task, the second item represents the cost of the logistics company to transport the logistics task to the cooperative distribution company, and the third item represents the distribution cost required by the cooperative distribution company. The objective function indicates that the points with $F > 0$ are selected, and the distribution points requiring transshipment are calculated; equation (2) indicates that the transfer point cannot transfer within the company; equation (3) indicates that only one company can be responsible for a demand task; equation (4) indicates that the distribution weight of the distribution company is not greater than the bearing weight of the company; equation (5) is an integer constraint.

(2) Establishment of path planning layer model:

Combined with the characteristics of logistics cooperative distribution path optimization under the logistics cloud service platform, the distribution cost in the distribution process is set as the optimization goal, where distribution cost = fixed cost + variable cost + Transshipment cost.

Objective function:

$$\min f = \sum_{n=1}^{M} \sum_{k=1}^{K} x_{0n}^{k} C + \sum_{m=1}^{M} \sum_{n=1}^{M} \sum_{k=1}^{K} x_{mn}^{k} c_{mn} d_{mn} + \sum_{p=1}^{G} \sum_{q=1}^{G} \sum_{k=1}^{K} \sum_{j=1}^{N} x_{pq}^{k} c_{pq} d_{pq} \times \operatorname{int} \frac{q_j}{Q}. \tag{6}$$

Constraints:

Distribution vehicle capacity constraints:

Table 2: Variable list 1.

Variable	Describe	Type
f	Objective function	Unknown
pq	Logistics company	Known
c_{pj}	Unit cost from logistics company P to logistics task J	Known
d_{pj}	Distance from logistics company P to task J	Known
c_j^p	Price from logistics company P to task J	Known
y_{jp}	Is the task j the responsibility of the logistics company P	0-1 variable

Table 3: List of variables 2.

Variable	Describe	Type
d_{mn}	The distance from distribution point m to n	Known
c_{mn}	Unit cost of transportation from distribution point m to n	Known
X_{mn}^k	The distribution demand of logistics distribution task J	Known
q_j	Distribution demand of task J	Known
H	H vehicles available	Known
Q	Maximum carrying capacity of vehicle	Known
q_j	The distribution demand of transshipment logistics distribution task J has been determined	Unknown

Table 4: Letters list.

Indexes	Meaning	Range
M, s, n	Node	1 to M
j	Task	1 to N
P, q	Company	1 to G
k	Type of shipping	1 to K

$$\sum_{n=1}^{M}\sum_{k=1}^{K} x_0^k, k, \tag{7}$$

$$\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{k=1}^{M} x_{mn}^k q_m^k \le Q, k = 1, 2, \ldots, K, \tag{8}$$

Distribution vehicle routing constraints:

$$\begin{gathered}\sum_{n=1}^{M} x_{0n}^k = 1,\\ k = 1, 2, \ldots, K,\end{gathered} \tag{9}$$

$$\begin{gathered}\sum_{m=1}^{M} x_{m0}^d = 1,\\ k = 1, 2, \ldots, K,\end{gathered} \tag{10}$$

$$x_{mnk}^d = 0, n = m; k = 1, 2, \ldots, K, \tag{11}$$

$$\sum_{m=1}^{M}\sum_{k=1}^{K} x_{mn}^k \le 1, n = 1, 2, \ldots, M, \tag{12}$$

$$\sum_{n=1}^{M}\sum_{m=1}^{M} x_{mn}^k = 1, k = 1, 2, \ldots, K, \tag{13}$$

$$\sum_{m=1}^{M} x_{ms}^{k} - \sum_{n=1}^{M} x_{sn}^{k} = 0, k = 1, 2, \ldots, K, \tag{14}$$

$$x_{mn}^{k} = \begin{cases} 1, \text{In distribution, the KTH car goes from node M to node N,} \\ 0, \text{In the distribution, the KTH car does not go from node M to node N.} \end{cases} \tag{15}$$

Among them, the objective function, namely, formula (6), represents the total cost of distribution activities, in which the first part represents the fixed transportation cost in logistics distribution activities, the second part represents the variable transportation cost after the logistics cloud service platform reprocesses and plans the distribution activities, and the third part represents the activity cost of distribution among different logistics enterprises. Int indicates rounding up, and intq_j/Q indicates vehicle transportation times; equations (7) and (8) represent the constraints on the number of vehicles and carrying capacity; equations (9) and (10) indicate that each delivery vehicle starts from the distribution point and returns to the distribution point after completion; equation (11) indicates that the distribution vehicle cannot drive from one demand point to its own demand point in each distribution process; equation (12) indicates that only one vehicle can be responsible for the distribution of demand points in each distribution; equation (13) indicates that in distribution, when the demand point has distribution demand, the demand point must have vehicle distribution; equation (14) ensures the connectivity of the route of the vehicle; equation (15) represents integer constraint.

4. Results and Analysis

4.1. Genetic Algorithm Design of Logistics Cooperative Distribution Model. In this paper, genetic algorithm is used to solve the logistics distribution path planning model. The detailed solution process is as follows.

4.1.1. Code. Based on the research of the basic principle and main program of genetic algorithm, combined with the characteristics of vehicle routing problem, this paper uses the coding method of natural number coding to design genetic operation.

When there are m logistics distribution points and the service vehicle is K, the chromosome structure can be expressed as

$$(0 - (0 - m_{11} - m_{12} - \ldots - m_{1s} - 0 - m_{21} - m_{22} - \ldots - m_{2s} - 0 - \ldots - m_{k1} - m_{k2} \ldots m_{ks} - 0), \tag{16}$$

where 0 means the distribution center, and each vehicle starts from the distribution center and needs to return to the distribution center, and m_{ks} means the k-th vehicle serves the s-th customer.

4.1.2. Determine the Number of Vehicles. When planning the route, we need to choose the number of vehicles reasonably. This paper uses the following formula to determine the number of vehicles:

$$k = \left\lfloor \frac{\sum_{j=1}^{M} q_j}{Q} \right\rfloor + 1, \tag{17}$$

where $\lfloor\rfloor$ means rounding down.

4.1.3. Fitness Function. The value of fitness function reflects the quality of individual performance. Properly convert the objective function formula (6) into formula (18), so as to calculate the value of fitness function.

$$F(x) = \frac{1}{f(x)}. \tag{18}$$

According to the above formula, when the distribution cost of the vehicle is smaller, the fitness function of the individual is higher; when the distribution cost of the vehicle is greater, the fitness function of the individual is lower.

4.2. Logistics Task Allocation Simulation. In order to verify the rationality of the logistics cloud task scheduling model based on genetic algorithm and the effectiveness of the algorithm, we carry out simulation experiments. In order to implement the simulation, we simplify the model and transform the complex multiobjective optimization problem into a single-objective optimization problem through preference vector. Fitness function value of genetic algorithm. The genetic algorithm and open-source algorithm on Python are used to simulate the model proposed in this paper. Assuming that there are 10 LCSPS, the response time of service application is shown

Table 5: Response time of LCSP service application (m).

LCSP	2	3	4	5	6
General service response time	4	3	7	5	6
LCSP number	7	8	9	10	11
General service response time	8	10	14	5	13

Table 6: Completion time of LCSD business (m).

85	83	10	70	93	45	18	9	7	35
96	65	88	89	61	60	55	61	38	21
97	29	36	36	73	13	56	88	51	100
40	40	99	16	67	14	50	49	51	97
59	29	70	67	57	52	66	50	70	25
86	79	61	34	77	22	42	62	40	82
94	32	13	37	90	12	52	99	76	100
58	64	64	80	60	52	87	82	65	7
38	28	12	49	9	76	17	23	16	97
22	50	47	98	59	81	69	92	88	71

Table 7: Server usage cost.

	LCSP1	LCSP2	LCSP3	LCSP4	LCSP5	LCSP6	LCSP7	LCSP8	LCSP9	LCSP10
Use cost	0.8	0.6	1.3	1.2	1	1.3	1.4	1.6	0.7	1.1

Table 8: Optimal dispatching scheme.

	LCSP1	LCSP2	LCSP3	LCSP4	LCSP5	LCSP6	LCSP7	LCSP8	LCSP9	LCSP10
Number of tasks	13	14	10	11	12	9	9	8	13	11
Resource utilization	74.4	78.3	71.2	76.6	75.4	73	73.9	72.2	77.1	72.8
Completion time	534	548	494	541	536	528	532	520	546	522

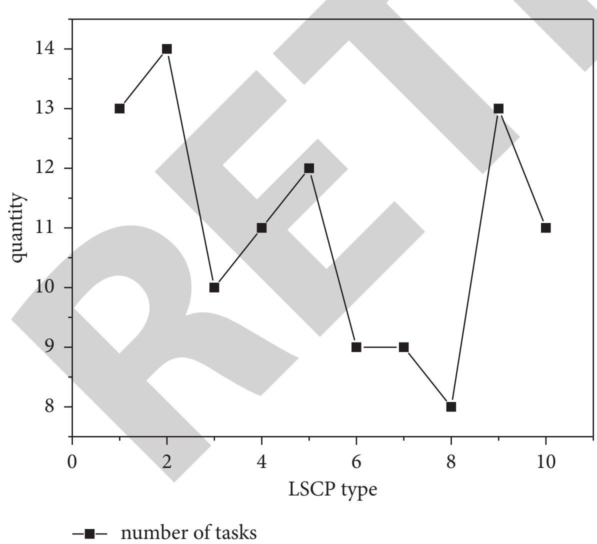


Figure 3: Number of LSCP scheduling tasks.

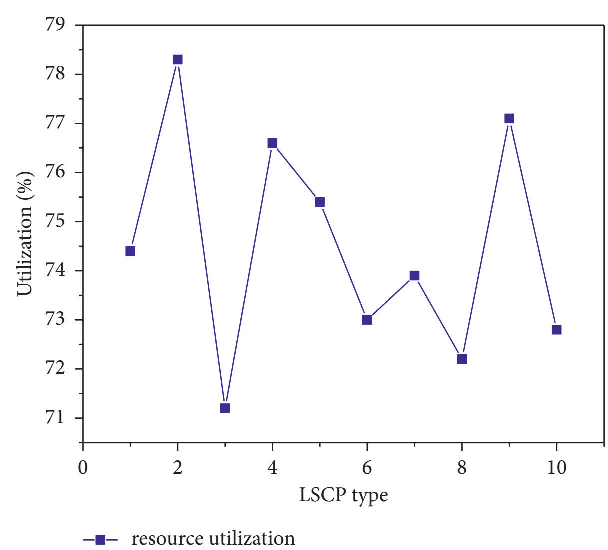


Figure 4: LSCP scheduling resource utilization.

in Table 5. The response time allocated to LCSPS is the fastest, so the response time allocated to LCSPS in different services is the shortest.

At the same time, there are 100 LCSD, and the business completion time is shown in Table 6. It can be seen that the completion time of these 100 tasks ranges from 6 minutes to 99 minutes, which means that once the allocation is improper, it will bring great time cost.

At the same time, the use cost of each server is shown in Table 7.

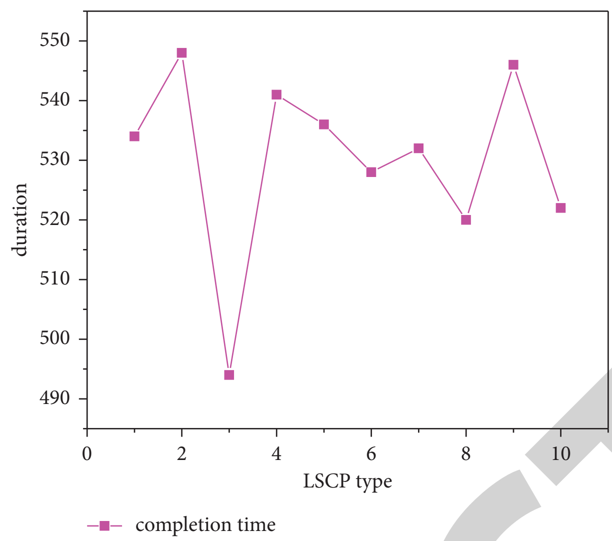


FIGURE 5: LSCP scheduling time.

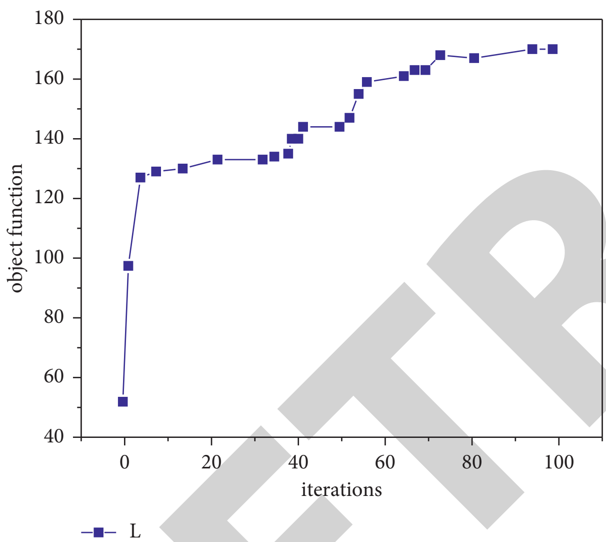


FIGURE 6: Variation curve of objective function value.

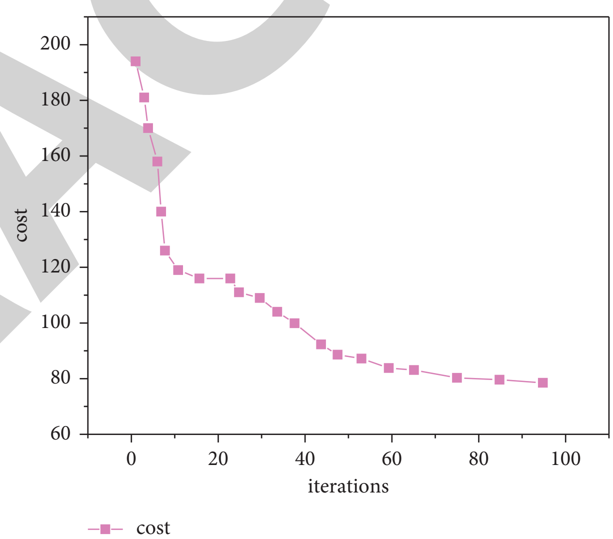


FIGURE 7: Change chart of dispatching cost.

Set the maximum service time of each LSCP to 12 hours per day, and select the preference vector as $\lambda = (0.4, 0.4, 0.2)$; that is, pay more attention to the scheduling time and resource utilization at this time. The optimal scheduling scheme is obtained through Python simulation, as shown in Table 8 and Figures 3–5.

Figure 6 shows the change curve of the objective function. It can be seen from the figure that after 100 generations of iterations the object function value increases rapidly from 30 to 130 and slowly from generation 5 to generations 40 to 130. Subsequently, the 40th generation to 60th generation are rapidly upgraded to 160. Finally, the 60th to 100th generations are basically stable at about 170.

Figure 7 shows the change of scheduling cost. It can be seen from the figure that the cost in the scheduling process gradually decreases with the increase of the number of algorithm iterations, from the initial unit cost of nearly 200 to 120. Then it gradually decreased to about 80. Genetic algorithm shows the ability of efficient and accurate solution in this 100-generation iteration.

Figure 8 shows the trend of total time change. It can be seen from the figure that the total time decreases rapidly with the number of iterations, from more than 1000 at the beginning to about 500 at last, and remains stable for a period of time, reducing by 50%. From the above simulation results, it can be seen that the fitness function value can converge after 100 generations of iterative calculation by using the genetic algorithm described in this paper. Our algorithm assigns 100 LCSD tasks to 10 LCSPS, and when the preference vector $\lambda = (0.4, 0.4, 0.2)$,

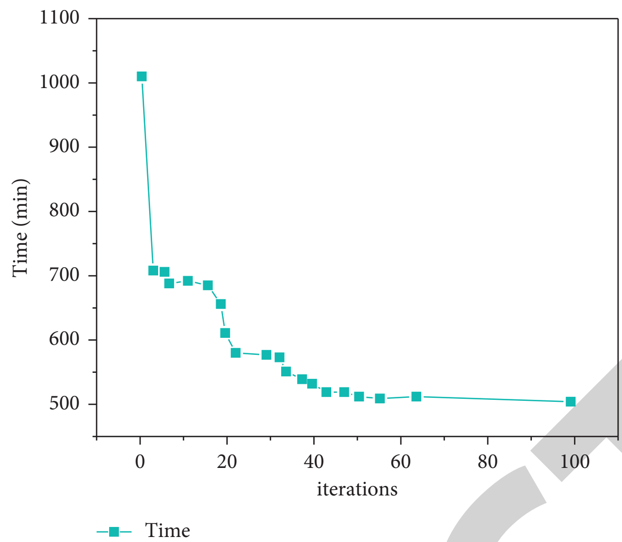


Figure 8: Total time variation trend.

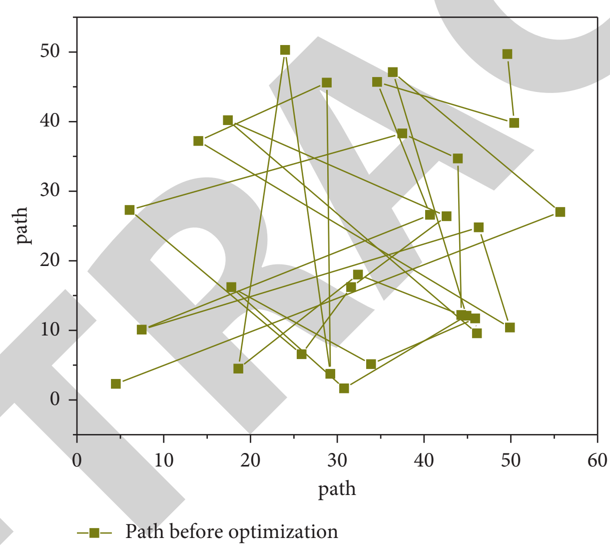


Figure 9: Schematic diagram of path before optimization.

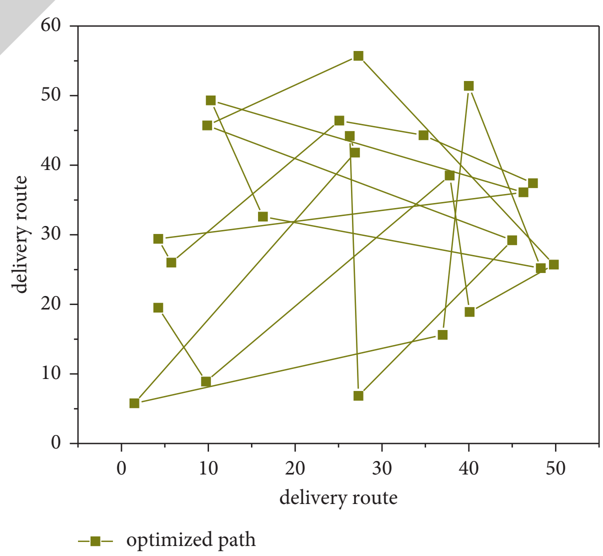


Figure 10: Schematic diagram of optimized distribution path.

one of the optimal solutions is 176.4. In this case, the service cost is 80.0 yuan and the service time is 507.2 minutes.

4.3. Logistics Cooperative Distribution Simulation. Suppose that the distribution center p of a logistics resource provider has a distribution area with 30 logistics distribution demand points at different locations, the weight of each task is no more than 2 kg, the maximum load weight of each distribution vehicle is 7 kg, the fixed cost of vehicles C is 2 yuan/vehicle, $C_{pj} = 1$ yuan/m, $C_{pq} = 1$ yuan/m, $C_{mn} = 1$ yuan/m, the number of vehicles $k_p = 6$, and the location of distribution center p is (25, 25). The location of the distribution center q of another logistics resource provider is (50, 40). There are 12 logistics distribution demand points in different locations. The weight of each task is no more than 2. The maximum load capacity of each distribution vehicle is 70 kg. The fixed cost of vehicles C is 2 yuan/vehicle, the number of vehicles $k_p = 7$, $C_{pq} = 2$ yuan/meter, $C_{pj} = 1$ yuan/meter, and $C_{mn} = 1$ yuan/meter. L_i^P represents the i-th distribution point of P company.

The distribution route before optimization using the algorithm in this paper is shown in Figure 9. At this time, the distribution cost is 2212.43 yuan, the distribution vehicles are 5, and the total mileage is 1894.46.

The genetic algorithm is used to solve the problem. The algorithm parameters are as follows: population size pop size = 300, maximum number of iterations max gen = 200, crossover probability PC = 0.8, and mutation probability PM = 0.1. Using the data of this paper and substituting it into the model established in this paper, the following distribution scheme is obtained: the simulation results are shown in Figure 10. It can be seen that the p minimum distribution cost is 601.58 yuan, the distribution vehicle is 5, and the total mileage is 477.41.

It can be seen that, after using the algorithm to optimize the path, the path interleaving is greatly reduced, and the vehicles do not take the repeated route, which can greatly save the cost. After calculation, the total mileage after optimization is 74.8% lower than that before optimization, and the cost is significantly reduced by 72.8%. To sum up, the last kilometer distribution algorithm proposed in this paper can greatly reduce the cost of logistics resource scheduling, which has obvious research significance.

5. Conclusion

In this paper, a logistics task allocation optimization model based on genetic algorithm is proposed. According to the logistics resource scheduling process and characteristics, the logistics resource scheduling process is divided into cloud logistics demand task allocation and the last kilometer distribution process of logistics goods; comprehensively consider multiple objectives: resource utilization, scheduling time, and operation cost, establish the logistics task allocation model, give the relevant decision variables, objective functions, and constraints, analyze and model the last kilometer problem of logistics distribution, and give the relevant decision variables, objective functions, and constraints, referring to the model established in this paper, combined with the relevant genetic algorithm to realize the effective design and solution, and through the python software to realize the simulation, so as to verify the feasibility and effectiveness of the model and algorithm established in this paper. The factors selected in this paper exclude many complex unpredictable and difficult to quantify situations, and the impact of practical problems has not been studied in a deeper level, including insufficient analysis of weather, terrain, road conditions, and so on. In the follow-up research, these factors should be gradually added to the whole discussion system, to reduce the error between the model built in the paper and the actual situation as much as possible and achieve the ideal optimization effect. At the same time, the application of existing intelligent technology in cargo transportation is just beginning, and there may be more advanced experience in other applications for reference. If these successful examples are absorbed, the effect of cargo transportation management may be better improved.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Shijiazhuang Institute of Railway Technology.

References

[1] Z. Li, Y. Li, W. Lu, and J. Huang, “Crowdsourcing logistics pricing optimization model based on DBSCAN clustering algorithm,” *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[2] S. Yang, “Optimization of urban logistics distribution path under dynamic traffic network,” *International Core Journal of Engineering*, vol. 6, no. 1, pp. 243–248, 2020.

[3] X. Jiang, J. Li, Y. Lu, and G. Tian, “Design of reverse logistics network for remanufacturing waste machine tools based on multi-objective gray wolf optimization algorithm,” vol. 8, no. 99, p. 1, IEEE Access, 2020.

[4] G. J. Adama, D. O. Jimoh, and M. Y. Otache, “Optimization of irrigation water allocation framework based on genetic algorithm approach,” *Journal of Water Resource and Protection*, vol. 12, no. 4, pp. 316–329, 2020.

[5] C. Lammie, J. K. Eshraghian, W. D. Lu, and M. R. Azghadi, “Memristive stochastic computing for deep learning parameter optimization,” *Circuits and Systems II: Express Briefs, IEEE Transactions on*, no. 99, p. 1, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9794561, 1 page
https://doi.org/10.1155/2023/9794561

Retraction

Retracted: Risk Assessment Method of Agricultural Management Investment Based on Genetic Neural Network

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zhao, "Risk Assessment Method of Agricultural Management Investment Based on Genetic Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 2373363, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2373363, 10 pages
https://doi.org/10.1155/2022/2373363

WILEY | Hindawi

Research Article

Risk Assessment Method of Agricultural Management Investment Based on Genetic Neural Network

JingJing Zhao

Changchun University of Finance and Economics, Jilin, Changchun 130122, China

Correspondence should be addressed to JingJing Zhao; xqb@nuc.edu.cn

Received 18 November 2021; Revised 22 December 2021; Accepted 21 January 2022; Published 29 April 2022

Academic Editor: Chin-Ling Chen

At present, the economy has entered the new normal. In the environment of economic slowdown, agricultural operation and investment risk management is of great significance in investment activities. In order to avoid the related risks of the agricultural operation and investment, improve the stability of agricultural operation and investment risk assessment, and ensure the effect of the agricultural operation and investment risk assessment, an agricultural operation and investment risk assessment method based on genetic neural network is proposed. This method fully considers the global optimization capability of the genetic algorithm and optimizes the weights and thresholds of the BP neural network. According to the construction principle of the index system, from the perspective of systemic risk and nonsystematic risk, a risk assessment index system for agricultural business investment has been established. Based on the establishment of risk evaluation index system, using the genetic neural network method, the agricultural operation investment risk evaluation model is established to realize the agricultural operation investment risk evaluation. The experimental results show that the proposed method has a good effect on the risk assessment of the agricultural operation and investment and can effectively improve the stability of the risk assessment of agricultural operation and investment.

1. Introduction

In the market economic environment, agricultural business enterprises are gradually emerging, which promote the development of the agricultural economy, and the real realization of modern agriculture also requires them to play a greater role [1–3]. Agricultural management enterprise is an enterprise integrating development, production, operation, and service. It can change the current situation of the economy and science and technology and organically combine economy and science and technology. Therefore, it plays a role of link [4, 5]. However, with the rapid economic development, the international competition of agricultural enterprises is becoming increasingly fierce, which increases various uncertain factors, and various risks faced by agricultural operation and investment are also complex. At present, the risk decision-making methods used by agricultural operations and investment managers for complex investment projects cannot accurately evaluate the size of their risks, so if they cannot carry out effective risk prevention, enterprises will inevitably suffer losses. Agricultural operation investment is a significant sign of great progress in agricultural development. Only with the good development of agricultural operation and investment, we can solve a series of problems related to farmers, such as employment pressure, low income, and rural urbanization, and stabilize the healthy development of other economic and related industrial sectors [6, 7]. Therefore, studying the risk management of agricultural operation and investment is of great and far-reaching significance to promoting the upgrading of the agricultural industry and reducing the project investment risk of agricultural enterprises.

At present, scholars in related fields have studied the risk assessment of operation and investment. Reference [8] proposes one belt, one road country risk assessment method for renewable energy investment under uncertainty. Using the TODIM method, they determine the risk factors of renewable energy investment and divide them into five categories:

economic, technological, environmental, social, and political. The fuzzy analysis network method is used to weight the identified factors, and the COPRAS, MABAC, and GRA methods are used to sort different renewable energy sources under uncertain conditions to realize the risk assessment of renewable energy investment. Finally, the sensitivity analysis of investment decision is carried out, focusing on the tradability of renewable energy projects. This method is helpful to evaluate investment projects and expand the international market. Reference [9] proposed the investment risk assessment of coal-fired power plants in the Belt and Road countries based on the ANP-Entropy-ToDIM method. An evaluation criteria system is established from eight dimensions, and the weights of these criteria are determined by the combinatorial analysis network process entropy method. Considering the psychological characteristics of decision-makers, the TODIM method is used to rank the overall risk level of CFPP investment in 23 countries to realize the investment risk evaluation of coal-fired power plants. This method has the maximum weight of the economic basis standard.

Based on the above analysis, this paper proposes a risk assessment method for agricultural operation and investment based on genetic neural network. Based on the analysis of the principle of BP neural network and genetic algorithm, and according to the construction principle of the index system, the risk evaluation index system of agricultural operation and investment is established. The genetic neural network method is used to construct the risk assessment model of agricultural operation and investment, so as to realize the risk assessment of agricultural operation and investment. Finally, it is verified that the proposed method has a better risk assessment effect and can effectively improve the stability of risk assessment.

2. Genetic Neural Network

2.1. Principle of BP Neural Network. BP (backpropagation) neural network takes the network error square as the objective function and uses the gradient descent method to calculate the minimum value of the objective function [10–12]. BP neural network has strong pattern recognition ability and nonlinear mapping ability, which is suitable for solving the problems of function approximation, pattern recognition, classification, and so on. The typical topology of the BP neural network is shown in Figure 1.

The information processing process of the BP neural network can be divided into two parts: forward propagation and back propagation.

(1) Forward propagation: By calculating the output of the previous layer of neurons as the input of the next layer of neurons, it is calculated by layer. For a labeled sample set $\{(x^i, y^i)\}_{i=1}^m$, where x^i, y^i represents the i training sample and label, respectively, and m is the sample dimension. Suppose the neural network model is $h_{w,b}(x)$, where w and b are the connection weights and biases between each layer of the model. Assuming that s^k is the input value of the k layer, the output of the $k+1$ layer neuron is expressed as follows:

$$s^{k+1} = w^k a^k + b^k, \quad (1)$$

$$a^{k+1} = f\left(s^{k+1}\right). \quad (2)$$

In formulas (1) and (2), w^k and b^k are, respectively, expressed as the weight and bias of layer k network, a^k is the output value of layer k neuron, and $f(\cdot)$ is the activation function of neuron, usually sigmoid function or tanh function [13–15]. The forward propagation of the network can be completed by successively calculating from the input layer to the output layer.

(2) Back propagation: Back propagation refers to the process of optimizing the loss function of the neural network with algorithms such as random gradient descent [16]. In order to make the learning results of the network model on the training samples more consistent with the label data, that is, to minimize the difference between the network output value and the label data, the loss function is defined as

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left\|\overline{y}^i - y^i\right\|_2^2 + \frac{\lambda}{2}\sum_i\sum_j w_{ij}^2. \quad (3)$$

The above formula consists of two parts. The first half represents the mean square error between the output value and label value of the network model on the training sample set. The latter half represents the weight attenuation term, which is used to reduce the weight amplitude value and prevent overfitting. In order to minimize formula (3), the following steps are usually carried out:

Step 1. The weight w and bias b are randomly initialized, the training samples are inputted into the network for calculation, and the output value is obtained.

Step 2. Optimization algorithms are used such as stochastic gradient descent to update w^k and b^k as follows:

$$w^k = w^k - \alpha\frac{\partial J(w,b)}{\partial w^k}, \quad (4)$$

$$b^k = b^k - \alpha\frac{\partial J(w,b)}{\partial b^k}. \quad (5)$$

In formulas (4) and (5), α is the learning rate, usually the value [0–0.1], adjusted according to the actual situation. It can be seen from the above formula that the key to the backpropagation algorithm is to calculate the partial derivative of the loss function $J(w,b)$ to the weight w and the bias b. Its partial derivative is expressed as

$$\frac{\partial J(w,b)}{\partial w^k} = \beta^{k+1}\left(a^l\right)^T, \quad (6)$$

$$\frac{\partial J(w,b)}{\partial B^k} = \beta^{k+1}. \quad (7)$$

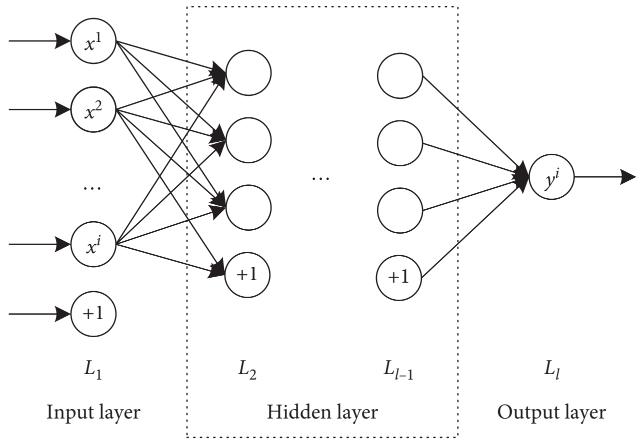


Figure 1: Typical topology of BP neural network.

In formulas (6) and (7), β^k is the residual value of neurons in the k layer of the network. On this basis, the difference between the output value of the output layer and the sample label is first calculated to obtain the residual value β^l of the output layer, then calculate the residual β^{l-1} of the previous layer in the reverse direction, and calculate all the residuals as follows:

$$\begin{aligned} \beta^l &= -\left(\overline{y}^i - y^i\right) f'\left(s^l\right), \\ \beta^k &= w^k \beta^{l+1} f'\left(s^k\right). \end{aligned} \tag{8}$$

Step 3. Iterate until the minimum value is obtained. The backpropagation algorithm searches the weight of the model in the high-dimensional space through the random gradient descent algorithm and continuously iterates the training process to minimize the loss function as much as possible. However, for multihidden layer networks, this algorithm often converges at the local minimum rather than the global minimum. The mathematical space of the error function of multilayer neural network is high dimensional, there are many gradient descent routes, and there are many routes including local extremum convergence. Random initialization of weights and offsets often leads to a loss function, resulting in falling into a local minimum of a certain dimension.

2.2. Principles of Genetic Algorithm. Based on the theory of evolution and genetic genetics, the genetic algorithm finds the optimal solution rather than the local solution of the actual problem by simulating the species selection and evolution in nature [17–19]. The optimization is based on the individual fitness obtained by the algorithm. The optimization method is mathematical iteration, and finally, the optimal solution of the problem to be solved is obtained. The specific process of the genetic algorithm is shown in Figure 2.

It can be seen from Figure 2 that the effect of the genetic algorithm in the actual optimization process is determined by the following five key factors:

(1) Coding of parameters to be optimized: Genetic algorithm first encodes the parameters of the problem; that is, the problem is encoded into a string. The simplest coding method is binary coding. The problem can be represented by a binary array. Another method is real number coding.

(2) Initial population setting: The population size will directly affect the optimization efficiency and running times of the genetic algorithm. The population size can be set according to the actual situation, and the general value range is [20, 200].

(3) Individual fitness calculation: Fitness is the standard for evaluating the quality of chromosome individuals in the population and the basis for selecting genetic operation [20–22]. In the specific application, the design of the fitness function should be combined with the requirements of solving the problem itself. In this paper,

$$f(E) = \frac{1}{1+E} \ (E \geq 0). \tag{9}$$

(4) Genetic optimization operator setting: Genetic operation can gradually optimize the solution of the problem and approach the optimal solution. Genetic manipulation includes selection, crossover, and variation [23–25]. Selection and crossover basically complete most of the search function of the genetic algorithm, and mutation increases the ability of the genetic algorithm to find near-optimal solution.

Selection: The selection process is the premise of cross mutation. The first step is to select the chromosomes that can produce the next generation population. The commonly used selection method is the roulette algorithm [26–28]. The probability of an individual being selected is positively correlated with the value of its fitness function. The probability of an individual $f(x_i)$ being selected can be expressed as follows:

$$f(x_i) = \frac{f(x_i)}{f(x_1) + f(x_2) + \cdots + f(x_n)}. \tag{10}$$

Crossover: Two chromosomes randomly A, B select a position to exchange some genes to produce two new chromosomes. The probability of chromosome crossing can be expressed by P_c.

Before chromosome crossing,

$$\begin{aligned} &A\text{: } 00000|011100000000|10000, \\ &B\text{: } 11100|000001111110|00101. \end{aligned} \tag{11}$$

After chromosome crossing,

$$\begin{aligned} &A'\text{: } 00000|000001111110|10000, \\ &B'\text{: } 11100|011100000000|00101. \end{aligned} \tag{12}$$

Variation: In the process of chromosome evolution, variation will occur with a certain probability, which is expressed by P_m.

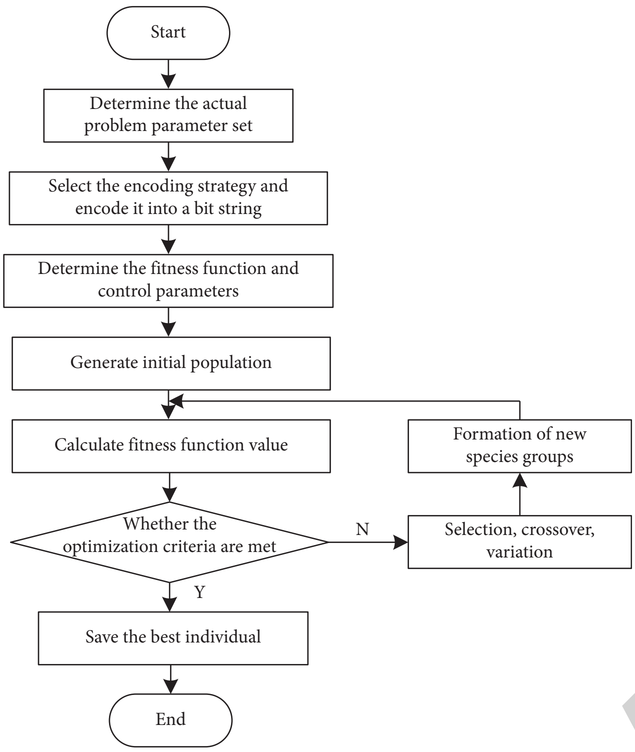


Figure 2: Specific process of genetic algorithm.

Before chromosome variation,

$$000001110000000010000. \quad (13)$$

After chromosome variation,

$$000001110000100010000. \quad (14)$$

(5) Genetic control operator setting: It mainly sets the algebra of genetic evolution. For the optimization with a small population number, the evolutionary algebra range should be [50, 100].

2.3. Construction of Genetic Neural Network. This paper uses a genetic algorithm to optimize the weights and thresholds of the BP neural network, constructs a genetic neural network, and achieves the purpose of obtaining the optimal solution [29, 30]. The basic principle of the genetic neural network is to solve it by a genetic algorithm. Because a genetic algorithm searches a group of points in the solution space at the same time and forms an evolving population sequence, some global good points can be obtained at the same time after evolving a certain algebra. Starting from these good points, the neural network is used to solve them, respectively, and then, the global optimization solution is obtained. The flowchart of the genetic neural network is shown in Figure 3.

The steps of using a genetic algorithm to assist in optimizing the weight of a neural network are as follows:

(1) A part from the total sample is randomly extracted, which is a set of weight thresholds, and effective coding methods are used to encode the group. A gene string represents a weight threshold distribution state of the network, and a chromosome represents a set of weight thresholds of the neural network.

(2) By calculating the error function of each chromosome in the neural network, the fitness function value of the population is further determined, and the constituent population in line with the target value is selected.

(3) The population with high fitness in line with the target is taken as the parent, and the next step is passed on to the offspring.

(4) To process and calculate the parents to get the next generation, crossover operator and mutation operator are used.

(5) Whether the obtained results meet the termination goal is determined. If it is satisfied, it will terminate and enter the next stage. If there are still deficiencies, it will repeat the cycle until satisfactory results are obtained.

3. Construction of a Risk Evaluation Index System for Agricultural Business Investment

3.1. Construction Principle of Index System. In order to fully consider the influencing factors of agricultural operation investment risk and strive to obtain scientific and objective evaluation results, the following principles should be observed when constructing the risk evaluation system:

(1) Scientific principle: The design of the risk evaluation index of agricultural operation investment must comply with relevant economic theories [31]. The selection of indicators should cover all risk factors as much as possible, and the concept and significance of indicators should be clear, which can accurately reflect the characteristics of the agricultural operation and investment risks, meet the requirements of risk evaluation, and ensure the comprehensiveness and objectivity of the results.

(2) Systematic principle: The risk assessment of agricultural operation and investment is a systematic problem. When selecting indicators, the indicators should be closely linked with the evaluation object, and the essential characteristics of the evaluation object should be systematically analyzed. The index system should not only reflect the overall logic but also highlight the hierarchy and relevance.

(3) Accuracy principle: Accuracy is reflected in two aspects: first, the evaluation index must accurately reflect the state of the evaluation object, and second, the index data must accurately meet the needs of the model; that is, the caliber of the data and the model must be consistent.

(4) Comparability principle: The design of indicators must comply with unified and general accounting

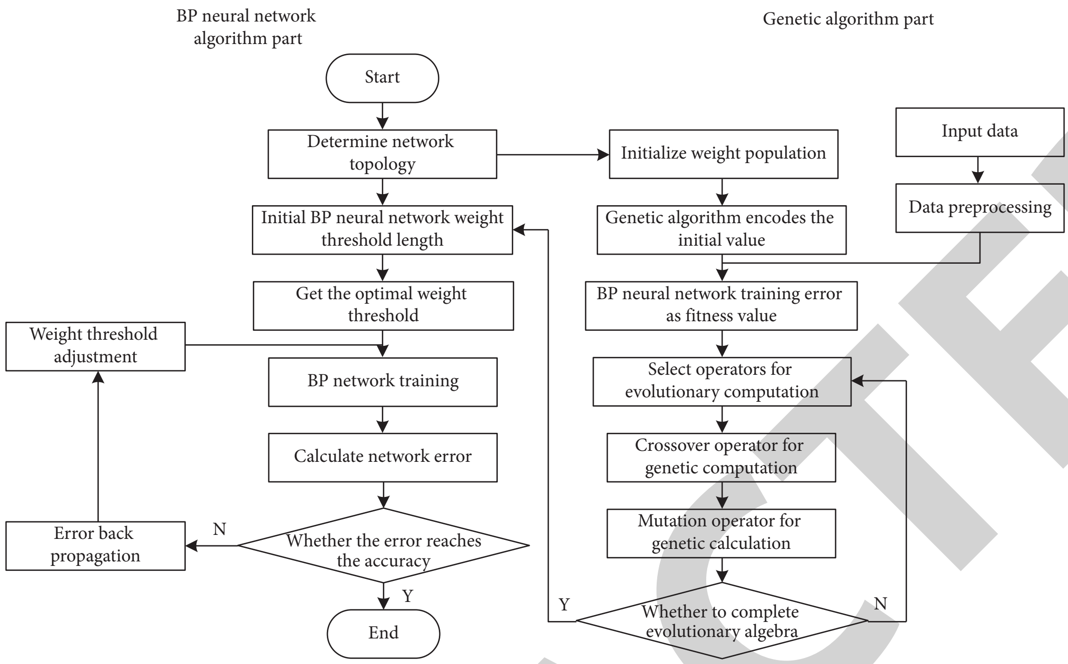


Figure 3: Flowchart of the genetic neural network.

methods and can be compared between different countries, regions, and projects. At the same time, attention should be paid to the measurement range and unit of measurement of indicator data, so as to facilitate horizontal or vertical comparison.

(5) Feasibility principle: The quantitative standard of indicators and the difficulty of obtaining relevant data shall be considered in the selection of indicators. The selected indicators shall be convenient for statistics, analysis, and acquisition. Indicators whose data are difficult to obtain or cannot be quantified may not be included in the index system for the time being.

3.2. Risk Evaluation Index System of Agricultural Operation Investment. Following the construction principle of the index system and integrating the opinions of technical authorities, senior experts, and agricultural risk investment experts are integrated. On this basis, systematic and nonsystematic risks are integrated to construct an index system for agricultural business investment risk assessment. Details are as follows.

Systemic risks mainly include the following:

(1) Policy and regulation risk: This risk refers to the unexpected negative impact on venture capital caused by imperfect policies and regulations or frequent adjustments and changes. Among them, the imperfect policies and regulations on venture capital are one of the most prominent potential risks.

(2) Macroeconomic fluctuation risk: This risk refers to the risk generated in the process of macroeconomic trend and fluctuation, such as economic recession, economic crisis, inflation, and deflation.

(3) Social factor risk: This risk refers to the investment risk caused by social unrest or social factors such as cultural customs, investment philosophy, and mass consumption tendency.

Nonsystematic risks mainly include the following:

(1) Technical risk: The advanced nature, maturity, practicality, matching, technological innovation, and scientific research talent resources are the key to winning the technological superiority and market advantage in the knowledge economy era and also the foundation for venture capital to gain high profits.

(2) Production risk: This risk refers to the risk caused by defects in production equipment status, organization system, operation and management mode, quality of production personnel, supply of raw materials, and intermediates.

(3) R & D risk: This risk is the core of the high-tech investment project and the key embodiment of the technical advantages. The R & D risk closely related to the project will affect each investment stage, and it is also the most direct embodiment of the project risk.

(4) Market risk: This risk mainly refers to the risk caused by the uncertainty of product market competitiveness, such as the degree of market demand and adverse factors in product substitutability, cost performance, marketing network, and life cycle, which may lead to market risk.

(5) Management risk: This risk refers to the risk caused by uncertain factors such as the quality and experience of managers, scientific decision-making,

rationality of enterprise management, selection of investment tools, size of investment scale, phased dynamic portfolio investment strategy, evaluation and screening of investment projects, capital structure of enterprises.

4. Construction of Risk Assessment Model of Agricultural Operation and Investment

4.1. Establish Network Topology. Based on the established risk evaluation system, according to the number of known risk measurement indicators, the number of different neurons in the input layer, hidden layer, and output layer is determined, so as to build a complete BP neural network structure model. The number of neurons in the input layer is the number of evaluation indexes, the result of the output layer is a single neuron, and the hidden layer can be obtained by model algorithm.

(1) Construct a set of risk evaluation factors: Let $U = \{u_1, u_2, \ldots, u_m\}$ be M risk indicators for evaluating agricultural investment. There are 21 evaluation indicators, and then, $M = 21$.

(2) Determination of decision rating: According to the management evaluation of agricultural operation and investment risk, the evaluation status of the agricultural operation and investment risk is obtained, as shown in Table 1.

(3) Sample training: According to the measurement indicators in the index system, data collection is combined with the actual qualitative data scoring situation, and the data are preprocessed to obtain the primary training sample set, sample set $M = \{m_1, m_2, \ldots, m_P\}$.

(4) Determination of the number of hidden layer neurons: According to the calculation formula of hidden layer neurons,

$$s = \sqrt{M + N} + \gamma. \tag{15}$$

In formula (15), M is the number of input neurons, N is the number of output results, and γ is a constant value between 0 and 1.

4.2. Genetic Algorithm Optimizes the Weights and Thresholds of Neural Networks. In terms of initial coefficient value selection and optimization, it is mainly composed of three parts: the threshold of hidden layer and output layer, the connection weight of input layer and hidden layer, and the connection weight of hidden layer and output layer [32, 33]. Each sample in the population set, as an individual, will reflect the ownership value and threshold between neurons in each layer of the BP neural network. The coding length of the sample individual for the genetic algorithm is equal to the sum of the number of thresholds and weights connected between all layers of the trained neural network of all samples in the population, which is

$$l = m \times n + s \times n + n. \tag{16}$$

In formula (16), s and m are the number of neurons in the hidden layer and input layer, respectively, and n is the number of neurons in the output layer. The weight matrix from the input layer to the hidden layer is as follows:

$$W = \left|w_{jr}\right|_{n\times s}. \tag{17}$$

The weight matrix from the hidden layer to the output layer is as follows:

$$V = \left|v_{rk}\right|_{s\times m}. \tag{18}$$

The threshold of hidden layer neurons is as follows:

$$B = \left|b_1, b_2, \ldots, b_s\right|. \tag{19}$$

The threshold of the output layer neuron is as follows:

$$C = \left|c_1, c_2, \ldots, c_m\right|. \tag{20}$$

The input vector of the genetic algorithm is as follows:

$$X = \left|w_{11}, w_{12}, \ldots, w_{13}, b_1, b_2, \ldots, b_s, v_{11}, v_{12}, \ldots, v_{13}, c_1, c_2, \ldots, c_m\right|. \tag{21}$$

4.3. Determination of Fitness Function. Assuming that P is selected to train the sample data, different training samples in the neural network get different training errors E^P. Based on the calculation basis of the sample output error square $(E^P)^2$, the total output error of the network for the overall training of the sample is

$$E_{\min}(k) = \frac{1}{2}\sum_{K=1}^{P}\left(d_k^p - o_k^p\right)^2, \quad p = 1, 2, \ldots, k. \tag{22}$$

In formula (22), o_k is the result vector of the output layer of the network, and d_k is the expected output vector obtained through sample training. According to the principle of model combination, the target search for the optimal solution is to select the weight threshold with the smallest error square and E performance by training in the neural network in the evolutionary generation of the sample. According to the evolution direction, the fitness function of the genetic algorithm can only increase continuously, so the fitness function combines the characteristics of the two as follows:

$$f\left(E_{\min}(k)\right) = \frac{1}{E_{\min}(k)}. \tag{23}$$

4.4. Parameter Setting of Genetic Operation. Based on the application principle of the genetic algorithm, the parameter value of the genetic operator in the genetic algorithm is set. According to the characteristics of the research subject and the core operating principle of the genetic algorithm, the parameters are effectively analyzed and set.

(1) Selection operator Φ: According to the selection strategy of the adaptation proportion of the fitness

Table 1: Risk assessment of agricultural operation investment.

Risk intervals	Risk status	Risk description
0–0.2	Safety	Network operation security
0.2–0.5	Relatively safe	Occasionally minor effects
0.5–0.67	Commonly	Cause small impact loss
0.67–0.85	More dangerous	Cause great difficulties
0.85–1	Danger	Causing serious losses or business interruption

value in the genetic algorithm, assuming that the fitness value of the i sample in the M sample individuals in the genetic algorithm is f_i, the selection probability of each sample individual i selected by the selection operator in the genetic operator is

$$\Phi_i = \frac{f_i}{\sum_i^M f_i}. \tag{24}$$

(2) Crossover operator P_r: Based on the principle comparison, the model in this paper chooses the arithmetic crossover method [34–36]. Arithmetic crossover assumes that Γ_A^t, Γ_B^t is the arithmetic crossover between two sample individuals. To execute the algorithm, it is necessary to determine the value of the coefficient α in a reasonable linear combination of each sample individual. If and only if α is a constant, the algorithm is uniform arithmetic crossover. An arithmetic crossover algorithm is a method to obtain new sample individuals by the pairwise linear combination of samples. Then, the new sample individuals are

$$\begin{aligned} X_A^{t+1} &= \alpha X_B^t + (1-\alpha) X_A^t, \\ X_B^{t+1} &= \alpha X_A^t + (1-\alpha) X_B^t. \end{aligned} \tag{25}$$

(3) Mutation operator P_Ψ: The genes carried by each sample individual are arranged on its chromosome, the corresponding gene is selected from the chromosome of the sample individual with the mutation probability Ψ, and the value is randomly replaced within the value range. If x_k is the mutation point of the gene, the value range is $|U_{\min}^k, U_{\max}^k|$, and h is randomly selected in accordance with [0, 1], and then,

$$x_k = U_{\min}^k + h \times \left(U_{\max}^k - U_{\min}^k\right). \tag{26}$$

(4) Selection of control parameters: The actual operation of the genetic algorithm needs to preset the following 4 algorithm operating parameters, which is T, M, P_r, P_Ψ. T represents the evolutionary algebra of the genetic algorithm, which is generally between 100 and 500 generations according to the research subject. M represents the size of the sample group, which is selected between 20 and 50 according to the sample size. P_r represents the crossover probability of the genetic algorithm, and the value ranges from [0.4–0.99]. P_Ψ represents the mutation probability of the genetic algorithm, and the value ranges from [0.0001–0.1].

4.5. Execute BP Algorithm. First, the genetic operation is performed. After continuous crossover, variation, heredity, and other generations of reproduction in the sample population, the target sample individual with the largest fitness is searched; that is, the initial weight and threshold of each layer connection of the optimized neural network are obtained, and the optimization results are assigned to the neural network to obtain the optimized neural network structure model and then train the sample data.

The data values of various risk measurement indicators of agricultural operation and investment are brought into the model for training. The data samples are used to learn the nonlinear function relationship, and then, the control samples are used to verify, and the final output value of agricultural operation and investment risk evaluation is obtained, which is the risk of agricultural operation and investment of the example. According to the grade range of each risk rating in the model, the risk grade of agricultural operation investment risk is determined by comparing it with the final agricultural operation investment risk value.

5. Experimental Analysis

5.1. Sample Data Acquisition and Analysis. In order to verify the effectiveness of the risk assessment method of agricultural operation and investment based on the genetic neural network, the project leader and some staff of an enterprise were selected to cooperate, and a total of 21 indicators and 23 groups of data were collected. Due to the different data types of the index data, the input data are processed for data standardization in order to be effective and reasonable. On this basis, a risk assessment model for agricultural business investment is established. In this paper, MATLAB is used for experimental analysis, and the BP network structure is determined according to known conditions. The number of input nodes corresponding to the number of risk indicators in the article is 21, and the number of output layer nodes is 1, which is the comprehensive risk value of the final risk evaluation of the method. According to the data sample size, the number of sample training is selected to be 2000 times, the training target is 0.0001, and the learning rate is 0.01. After experimental adjustments, the initial population size is set $M = 50$, then the encoding method is selected to encode the weight threshold, encoding length $l = 18 \times 9 + 9 \times 1 + 9 + 1 = 180$, genetic algebra $T = 150$, crossover probability $P_r = 0.7$, mutation probability $P_\Psi = 0.05$, and the number of iterations is 100.

5.2. Effect Analysis of Risk Assessment of Agricultural Operation Investment. After genetic optimization, according to the conditions that the optimal sample individual meets, the sample individual with the best fitness is searched, and the optimal decoding value is obtained and applied to the established neural network structure to provide the initial weight threshold between the connections of each layer of the network and train. After the network training, any part of the samples is selected to predict their risk decision-making comments by the network, and regression fit the other training labels. The training fitting of the genetic neural network is shown in Figure 4.

It can be seen from Figure 4 that the training label corresponding to the training sample is basically consistent with the prediction risk decision-making comments obtained by the trained model judging the training data. It can be seen that the genetic neural network risk assessment effect obtained by the training of the proposed method is good and can be used for agricultural operation and investment risk assessment.

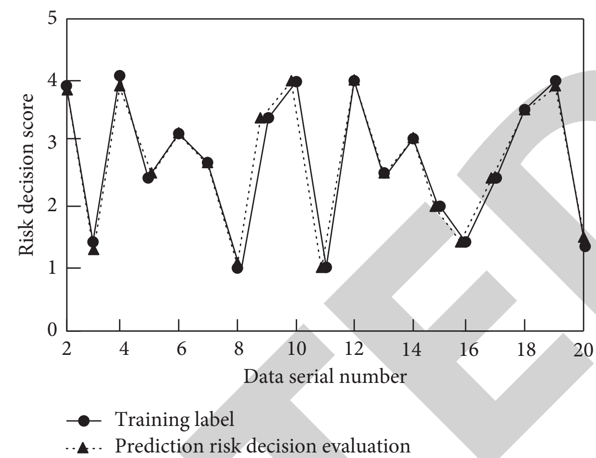


Figure 4: Training output results of the proposed method.

5.3. Stability Analysis of Agricultural Management Investment Risk Assessment. In order to further verify the stability of the agricultural operation investment risk assessment of the proposed method, the correlation coefficient δ is used to carry out the stability analysis of the agricultural operation investment risk assessment. The higher the value of δ, and the higher the overall stability of the risk assessment. The calculation formula is as follows:

$$\delta = \frac{\sum_{i=1}^{\sigma} (y(i) - \overline{y})\left(y'(i) - \overline{y}\prime\right)}{\sqrt{\sum_{i=1}^{\sigma} (y(i) - \overline{y})^2 \sum_{i=1}^{\sigma} \left(y'(i) - \overline{y}\prime\right)^2}}. \quad (27)$$

In formula (27), σ is the number of test samples, $y(i)$ is the test label, $y\prime(i)$ is the prediction result, $\overline{y}$ is the average value of the test labels, and $\overline{y}$' is the average value of the prediction results. Through the above experiment, according to formula (27), the fit between the prediction result of the proposed method and the label of the test sample is calculated as shown in Figure 5.

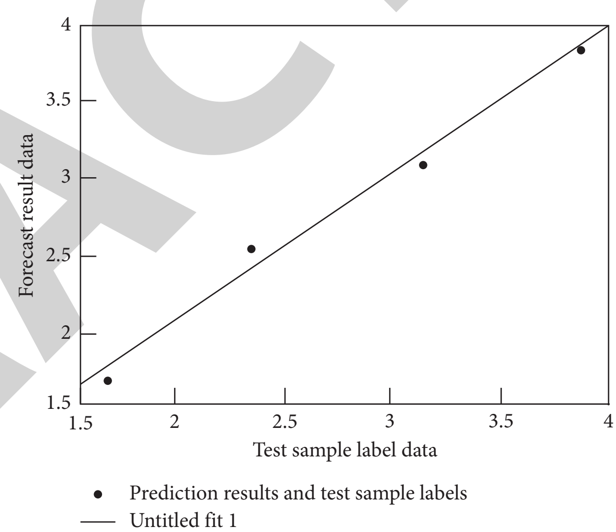


Figure 5: Fit between the prediction results of the proposed method and the labels of test samples.

It can be seen from Figure 5 that the prediction result has a linear relationship with the test sample label. The correlation coefficient $\delta = 0.9908$ is calculated between the two. It can be seen that the agricultural management investment risk assessment model of the proposed method has high stability.

5.4. Analysis of Key Indicators for Risk Assessment of Agricultural Operation Investment. The risk evaluation of agricultural operation investment is of great significance to the actual agricultural operation. In order to enable managers to better grasp the focus of work, the experiment uses the increase and decrease component method as an analysis tool to measure the key degree of various indicators to the risk of agricultural operation and investment. The calculation formula is as follows:

$$\omega = \frac{1}{n}\left[\sum_{j>i} (p(i,j) - p(i+1,j)) + \sum_{j>i} (p(j,i) - p(j,i-1))\right]. \quad (28)$$

In formula (28), $p(i,j)$ represents the evaluation relevance obtained from the i to j risk evaluation index which is used in the evaluation of agricultural business investment risk. ω represents the contribution of the i risk evaluation index to the risk evaluation of agricultural business investment. Among them, the larger the ω, the greater the contribution of the indicator, that is, the more critical the indicator. Through the above experiments, the contribution distribution of indicators in the risk assessment system of the proposed method is calculated according to formula (28), as shown in Figure 6.

It can be seen from Figure 6 that in the whole risk evaluation index system, the production risk index is the most important for the risk evaluation of the whole agricultural operation and investment and should be the focus of agricultural operation and investment. In the production risk, the operation and management mode should be taken as the core evaluation index, and the rest of the work should be carried out together. The management risk index is of

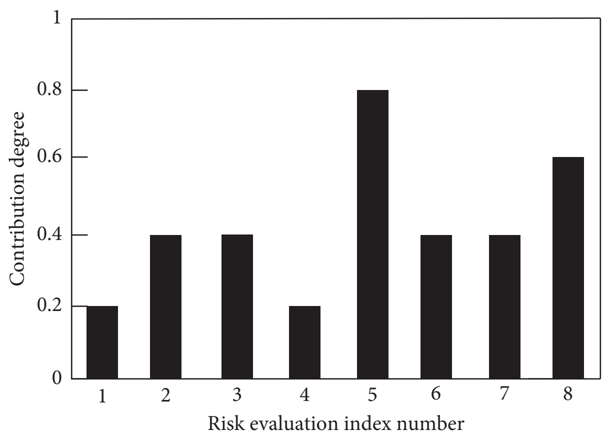


Figure 6: Contribution degree distribution of indicators in the risk evaluation system of the proposed method.

secondary importance to the risk evaluation of agricultural operation and investment and should be paid more attention to after the production risk. The contribution of each subindex in risk management is equal, and equal attention should be paid to it.

From the above analysis, it can be seen that the BP neural network optimized by genetic algorithm in this method evaluates the risk assessment of agricultural operation and investment, the fitting degree of the output results is high, and the assessment effect is good. Therefore, this method can be used as an effective technical means to determine the risk of agricultural operation and investment. At the same time, the overall stability of this method is high, and the contribution of each index can be determined accurately.

6. Conclusion

The risk assessment of agricultural operation investment proposed in this paper has a good effect and has high stability in the risk assessment of agricultural operation and investment. In the risk assessment index system, production risk and management risk have a great impact on the whole evaluation results, which should be the focus of agricultural operation and investment management. However, this method only considers the actual situation of an enterprise and a single enterprise in the process of the agricultural operation and investment risk assessment. Therefore, in the next research, we need to study the risk evaluation of agricultural operation and investment from the perspective of multienterprise cases, there is a large amount of enterprise data support, and the effect of the risk evaluation model will be better.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] X. Xi and J. Zhang, "Complexity analysis of a decision-making game concerning governments and heterogeneous agricultural enterprises with bounded rationality," *Chaos, Solitons & Fractals*, vol. 140, no. 11, Article ID 110220, 2020.

[2] O. Bezpartochna, "Forecasting the state of agricultural enterprises based on the results of economic diagnostics," *Journal Of Scientific Papers*, vol. 6, no. 1, pp. 3–11, 2021.

[3] D. C. Monckton, "Economic impacts of coal seam water for agricultural enterprises, lessons for efficient water management," *Sustainable Water Resources Management*, vol. 5, no. 1, pp. 333–346, 2019.

[4] H. Godlewska-Majkowska and A. Komor, "Energy as a factor of investment attractiveness of regions for agricultural enterprises," *Energies*, vol. 14, no. 9, pp. 2731–2738, 2021.

[5] O. Kovalenko, L. Yashchenko, and V. Zhuk, "Diagnostics of the state of agricultural enterprises financial security," *Accounting and Finance*, vol. 1, pp. 52–60, 2021.

[6] Q. Huang, Y. Xu, and C. Courcoubetis, "Financial incentives for joint storage planning and operation in energy and regulation markets," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3326–3339, 2019.

[7] M. F. Bellemare, K. Chua, J. Santamaria, and K. Vu, "Tenurial security and agricultural investment: evidence from Vietnam," *Food Policy*, vol. 94, no. 7, Article ID 101839, 2020.

[8] A. Hashemizadeh, Y. Ju, B. Smhb, and H. Phong, "Renewable energy investment risk assessment in belt and road initiative countries under uncertainty conditions," *Energy*, vol. 214, no. 1, Article ID 118923, 2020.

[9] J. Yuan, X. Li, C. Xu, C. Zhao, and Y. Liu, "Investment risk assessment of coal-fired power plants in countries along the Belt and Road initiative based on ANP-Entropy-TODIM method," *Energy*, vol. 176, no. 1, pp. 623–640, 2019.

[10] S. Tang and F. Yu, "Construction and verification of retinal vessel segmentation algorithm for color fundus image under BP neural network model," *The Journal of Supercomputing*, vol. 77, no. 1, pp. 1–15, 2021.

[11] M. Tsukada, M. Kondo, and H. Matsutani, "A neural network-based on-device learning anomaly detector for edge devices," *IEEE Transactions on Computers*, vol. 69, no. 7, pp. 1027–1044, 2020.

[12] Z. Liu, "Construction and verification of color fundus image retinal vessels segmentation algorithm under BP neural network," *The Journal of Supercomputing*, vol. 77, no. 7, pp. 7171–7183, 2021.

[13] J. Liu, X. Yang, Y. Feng, and P. Cui, "New fractional derivative with sigmoid function as the kernel and its models," *Chinese Journal of Physics*, vol. 68, no. 12, pp. 533–541, 2020.

[14] M. Tripathy, S. R. Chiluveru, and B. Mohapatra, "An accuracy controlled iterative method for efficient sigmoid function approximation," *Electronics Letters*, vol. 56, no. 18, pp. 854–859, 2020.

[15] A. Darwish, H. M. Ahmed, A. H. Arnous, and M. F. Shehab, "Optical solitons of Biswas–Arshed equation in birefringent fibers using improved modified extended tanh-function method," *Optik - International Journal for Light and Electron Optics*, vol. 227, no. 23, Article ID 165378, 2020.

[16] E. Yan, J. Song, C. Liu, J. Luan, and W. Hong, "Comparison of support vector machine, back propagation neural network and extreme learning machine for syndrome element differentiation," *Artificial Intelligence Review*, vol. 53, no. 4, pp. 2453–2481, 2020.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9824571, 1 page
https://doi.org/10.1155/2023/9824571

Retraction

Retracted: Selective Partial Update Adaptive Filtering Algorithms for Block-Sparse System Identification

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] D. Wei and Q. Xu, "Selective Partial Update Adaptive Filtering Algorithms for Block-Sparse System Identification," *Security and Communication Networks*, vol. 2022, Article ID 2207115, 7 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2207115, 7 pages
https://doi.org/10.1155/2022/2207115

Research Article

Selective Partial Update Adaptive Filtering Algorithms for Block-Sparse System Identification

Dandan Wei and Qing Xu

School of Information Engineering, Zunyi Normal College, Zunyi, China

Correspondence should be addressed to Dandan Wei; monicawdd1991@gmail.com

Received 21 February 2022; Revised 19 March 2022; Accepted 28 March 2022; Published 29 April 2022

Academic Editor: Chin-Ling Chen

The block-sparse normalized least mean square (BS-NLMS) algorithm which takes advantage of sparsity, successfully shows fast convergence in adaptive block-sparse system identification, adaptive control, and other industrial informatics applications. It is also attractive in acoustic processing where long impulse response, highly correlated and sparse echo path are encountered. However, the major drawback of BS-NLMS is largely computational complexity. This paper proposes a novel selective partial-update block-sparse normalized least mean square (SPU-BS-NLMS) algorithm. Compared with conventional BS-NLMS for block-sparse system identification, the proposed elective partial-update block-sparse NLMS algorithm takes partial-update blocks scheme which is determined by the smallest squared Euclidean-norm at each iteration instead of entire block coefficients to save computations. Computational complexity analysis is conducted to help researchers select appropriate parameters for practical realizations and applications. Computer simulations on acoustic echo cancellation are conducted to verify the results and the effectiveness of the proposed algorithm.

1. Introduction

System identification is frequently encountered in many applications such as acoustic echo cancellation (AEC) [1], interference suppression in industrial [2], and biomedical engineering [3]. In order to adapt to the time-varying characteristics of the statistical speech signal, the LMS algorithm or the normalized form is usually used to iteratively identify the unknown system [4]. In many scenarios, like network echo cancellation, the impulse responses are sparse which means most of the tap-weights are zero or small value, and there are few nonzero or large coefficients. There are several different kinds of sparse systems. The typical one such as TV transmission channels [5] is called a block-sparse system or a block-compressible system. Different from general impulse response sparse systems in which large coefficients are randomly distributed, the nonzero coefficients of block-sparse systems are composed of one or more clusters, and a cluster is a set of nonzero or large coefficients [6]. Multichannel input and multichannel output and satellite communication communications in confined spaces are a representative example of a multiclustering sparse system. Unlike multicluster-like sparse systems, the unit impulse response of acoustic echo channels is a single-cluster-like sparse system consisting of a long tail delay and an active part [7, 8].

Considering the speech echo path is a representative single-clustering sparse system, there is only one gathering of non-zero coefficients [9]. Several methods based on this priori knowledge have been presented including the group-zero-attracting LMS (GZA-LMS) algorithm and its improved version, the sparsity constraint LMS algorithm [10], the block-sparse LMS (BS-LMS) algorithm [11] and its input normalization variant version, the BS-NLMS algorithm, the block-sparse proportionate normalized LMS (BS-PNLMS) [10], and the BS-IPNLMS [12]. Compared with traditional LMS algorithms, the BS-LMS algorithm inserts a group of zero attraction of adaptive tap-weights into the cost function. Then, the optimization of group zero attraction of adaptive tap-weights is introduced to the block-sparse proportionate NLMS. Motivated by this, the computation reduction of the BS-LMS algorithm per iteration will be discussed in this paper by partial updating. Partial update is an efficient technique to reduce computations. It is very attractive for practical realizations and applications in hardware and digital signal processors. The representative algorithms include the preupdate strategy and the continuous NLMS algorithm [13], the M-max NLMS [14] algorithm that does not require extra operations to determine the value of tap-weights, and the set-membership PU-NLMS algorithm [15] that adopts the nonfixed filter coefficient update strategy.

In this paper, we will broadly categorized partial-updating adaptive algorithms into two classes based on different updating strategies as follows: first, various kinds of using certain data-dependent selection criteria adaptive algorithms were proposed to update the tap-weights, including the selective-block-update NLMS (SBU-NLMS) algorithm [16], the M-max NLMS algorithm (Max-NLMS) [17], the set-membership PU-NLMS algorithm [18]and its improved version, the L-norm-based algorithm [19]. These algorithms generally use the characteristics of unfixed updating strategies to achieve a faster convergence rate, like the BS-LMS and GZA-LMS. However, the drawback of certain data-dependent selection criteria adaptive algorithms is lower convergence rate for nonstationary signals because of data-dependent updating [18]. The second type is predetermined updating schemes adaptive algorithms [20], including the periodic LMS algorithm [21], the sequential PU LMS algorithm [22] which updates filter coefficients periodically, and a novel stochastic PU LMS (SPU-LMS) algorithm. These algorithms aim to lower the steady-state error of adaptive filters. However, the drawbacks of these algorithms are that they have to set many parameters. Furthermore, a novel variant of the S-LM algorithm was proposed to improve the robustness of this algorithm [23]. The sequential block LMS (SB-LMS) algorithm and its normalized version [23] for acoustic echo cancellation problems and the sequential block NLMS (SB-NLMS) algorithm [24] were proposed to speed up the convergence rate.

We proposed a novel SPU-BS-NLMS algorithm. The work focuses on the loss of computational complexity by selective partial update of blocks which chooses the smallest squared-Euclidean-distance as an optimization criterion in this paper. The selection criterion can be divided into two classes. The first class can update the blocks of the tap-weights vector by using squared Euclidean norms. The second class of algorithms can use larger gradient vector components. The complexity reduction of the SPU-BS-NLMS algorithm is achieved by dividing the filter taps into blocks and iterating only one block at a time instead of all the filter taps. As an indispensable part of AEC, the voice activity detection (VAD) algorithm technique of the speech signal distinguish the active/inactive speech periods will be utilized in this paper.

In Section 2, the BS-NLMS algorithm and voice activity detection technique are briefly reviewed. The proposed SPU-BS-NLMS algorithm and its computational complexity are derived and analyzed in Section 3. Computer simulations on acoustic echo cancellation in automobiles and tracking performance are conducted in Section 4. Notation: the notation $(.)^T$ is used for transpose, and $E(.)$ for taking expectation.

2. BS-NLMS Algorithm and VAD

2.1. BS-NLMS Algorithm. The system to be identified of the acoustic echo cancellation system is shown in Figure 1. Acoustic echo cancellation (AEC) aims to use the reference signal to cancel the echo in the microphone signal. The best goal of the echo cancellation algorithm is to achieve zero echo leakage and no distortion of the target speech. Of course, since noise is also unavoidable in the acoustic environment, it is sometimes necessary to take into account noise processing. The far-end user sends out a speech signal $u = [u_1(n), u_2(n), \ldots, u_L(n)]$. The speech signal passes through the unknown echo path to get the desired signal $d_n = u_n^T s + v_n$, where L denotes the length of echo path, and v_n denotes the additive background noise. The background noise is assumed to follow Gaussian distribution with zero mean. The core of echo cancellation is to use the LMS algorithm to find out w to model the echo path. w is the estimation of unknown system. The error signal e_n is expressed as

$$e_n = d_n - u_n^T w_n. \tag{1}$$

The smaller the error signal value is, the closer the estimated echo value is to the real echo. According to [12], the cost function of the least mean square algorithm can be expressed as $\xi_n = E[e_n^2]$. Then, using the negative gradient steepest descent method of the mean square error based algorithm, namely, the well-known least mean square adaptive algorithm, is expressed as

$$w_{n+1} = w_n + \mu e_n u_n, \tag{2}$$

where μ is the step size to adjust the balance between steady-state error and convergence speed of the adaptive algorithm. The low-order mixed $L_{2,0}$ norm of vector $w = [w_1(n), w_2(n), \ldots, w_L(n)]^T$ is the zero attractive force exerted by improving the block-sparse characteristic of the echo channel. The expression is as follows:

$$\|w\|_{2,0} = \left\| \begin{bmatrix} w_{[1]\|2} \\ w_{[2]\|2} \\ \cdots \\ w_{[N]\|2} \end{bmatrix} \right\|_0, \tag{3}$$

where N is the number of groups of echo canceller orders, and $w_{[j]} = [w_{(j-1)P+1}, w_{(j-1)P+2}, \ldots, w_{jP}]$ denotes the jth group of w. P denotes the group partition and can always divide evenly L. According to [12], the cost function of BS-NLMS algorithm can be rewritten as follows:

$$\xi_n = E\left[e_n^2\right] + \lambda\|w\|_{2,0}. \tag{4}$$

This cost function is a quadratic convex function with respect to the tap coefficients, so there must be a globally

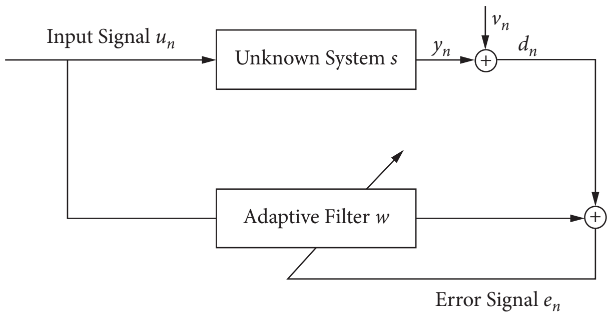


FIGURE 1: Block diagram for unknown system identification scenario.

unique minimum, and it is a positive factor. Then, using the negative gradient steepest descent method which is a similar derivation as that for NLMS, to obtain the coefficient update formula as follows:

$$w_{n+1} = w_n + \mu e_n u_n + \kappa g(w_n), \tag{5}$$

where $g(w)$ is expressed as formula (6), $\kappa = \mu * \lambda/2$ is used to adjust the intensity of block-sparse penalty and is step size parameter.

$$g(n) \triangleq [g_1(n), g_2(n), \ldots, g_L(n)]^T, \tag{6}$$

$$g_i(w) = \begin{cases} 2\alpha^2 w_i - \dfrac{2\alpha w_i}{\|w[\lceil i/p \rceil]\|}, & 0 < \left\| w\left[\lceil \dfrac{i}{p} \rceil\right] \right\| \le \dfrac{1}{\eta}; \\ 0, & \text{elsewhere.} \end{cases} \tag{7}$$

where symbol $\lceil \bullet \rceil$ denotes ceiling function, and η is a positive constant. For practical realization, a particularly small positive number will be inserted into the denominator to avoid division by zero. According to the description of the BS-NLMS algorithm in the second part of the article, the computational complexity is $O(L)$. In real applications, partial update provides an effective solution for saving computational complexity because the method only needs to update some parameters instead of all parameters at each iteration.

2.2. VAD Scheme. As an indispensable part of AEC, the voice activity detection (VAD) algorithm technique of the speech signal distinguish the active/inactive speech periods will be utilized in this paper. The voice activation detection algorithm extracts speech feature parameters and then needs to select specific decision criteria according to the application of the VAD detector to obtain the detection result. Since the third-order statistic of Gaussian noise is zero, for the speech signal whose environmental noise is Gaussian noise, there is a big difference between them. If a set of random signal $u_1, u_2, \ldots u_n$ is a real stationary discrete time signal, then r th-order joint cumulant is

$$C_{k_1 \ldots k_n} = (-j)^r \frac{\alpha^r \ln\phi(\omega_1, \omega_2, \ldots, \omega_n)}{\partial\omega_1^{k_1} \partial\omega_2^{k_2} \ldots \partial\omega_n^{k_n}} \Big|_{\omega_1 = \omega_2 = \ldots = \omega_n = 0}, \tag{8}$$

where $\phi(\omega_1, \omega_2, \ldots, \omega_n) = E[\exp\ j(\omega_1 u_1 + \omega_2 u_2 + \cdots + \omega_n u_n)]$ represents the k th-order moment function of the stationary signal. Then, the relationship between the moments τ and cumulant sequence are as follows:

Second-order cumulant

$$C_{2u}(\tau_1) = E[u(n)u(n + \tau_1)]. \tag{9}$$

Third-order cumulant

$$C_{3u}(\tau_1, \tau_2) = E[u(n)u(n + \tau_1)u(n + \tau_2)]. \tag{10}$$

Fourth-order cumulant

$$\begin{aligned} C_{4u}(\tau_1, \tau_2, \tau_3) &= E[u(n)u(n + \tau_1)u(n + \tau_2)u(n + \tau_2)] - C_{2u}(\tau_1), \\ &\quad C_{2u}(\tau_2 - \tau_3) - C_{2u}(\tau_2)C_{2u}(\tau_3 - \tau_1) - C_{2u}(\tau_3)C_{2u}(\tau_1 - \tau_2). \end{aligned} \tag{11}$$

Furthermore, by letting in $\tau_1 = \tau_2 = \tau_3 = 0$, we obtain, respectively, the variance γ_{2u}, skewness γ_{3u}, and kurtosis γ_{4u} as follows: variance $\gamma_{2u} = C_{2u}(0) = E[y^2(n)] = \sigma^2$, skewness $\gamma_{3u} = C_{3u}(0, 0) = E[y^3(n)]$, and kurtosis. $\gamma_{4u} = C_{4u}(0, 0, 0) = E[y^4(n)] - 3\gamma_{2u} = E[y^4(n)] - 3\sigma^2$.

3. Proposed SPU-BS-NLMS Algorithm

The voice activation detection algorithm extracts speech feature parameters at first, and then needs to select specific decision criteria according to the application of the VAD detector to obtain the detection result. Since the third-order statistic of Gaussian noise is zero, for a speech signal whose environmental noise is Gaussian noise, there is a big difference between the third-order statistics of the signal frame and the noise frame, and it is easy to judge the speech segment and the noise segment. Therefore, the third-order statistic can be used to detect the activation of speech. The output result of the VAD algorithm is represented by a Boolean value. When it is judged as a speech frame, the result is "1"; when it is judged as background noise, the result is "0." The fundamental partial-update scheme of the SPU-BS-NLMS algorithm is to update only some of the tap-weights per iteration instead of all by block selection criteria. It should be assumed that P and C can always divide L because the BS-NLMS algorithm already consider the filter coefficients into equal group partition which is different from several tradition methods in literature. C denotes the partition number of filter blocks and B is the length of coefficient block. The coefficients vector and input signal vector can be expressed as follows:

$$u(n) = [u_1^T(n), u_2^T(n), \ldots, u_C^T(n)], \tag{12}$$

$$w(n) = [\, w_1^T(n)\ \ w_2^T(n), \ldots, w_C^T(n) \,], \tag{13}$$

with

$$u_1(n) = [\, u_1^T(n)\ \ u_2^T(n), \ldots, u_B^T(n) \,], \tag{14}$$

and

$$w_1(n) = [\, w_1^T(n)\ \ w_2^T(n), \ldots, w_B^T(n) \,]. \tag{15}$$

The posterior and prior error vectors are $e_i(n+1) = d_i(n) - u_i(n)^T w_i(n+1)$ and $e_i(n) = d_n(n) - u_i(n)^T w_i(n)$, respectively. For a given updated block denoted i, the SPU-BS-NLMS algorithm is obtained by following optimisation criteria which is the minimized $L_{2,0}$-norm of error vector with a constraint.

$$\min \left\| w_i(n+1) - w_i(n) \right\|_{2,0}, \tag{16}$$

$$\text{Subject to}\, d(n) = w^T(n+1)u(n). \tag{17}$$

Combining equations (13) and (14), the cost function of SPU-BS-NLMS can be derived by using the method of the Lagrange multiplier.

$$\xi_i(n+1) = E\left[e_n^2\right] - \beta \left\| w_i(n+1) - w_i(n) \right\|_{2,0}, \tag{18}$$

where β is a Lagrange multiplier. The update formula is obtained by setting the partial derivative of the cost function to the weight vector equal to zero. The update equation is expressed as follows:

$$w_i(n+1) = w_i(n) + \frac{\mu e(n) u_i}{\left\| u_i \right\|^2 + \delta} + \kappa g(w_i(n)). \tag{19}$$

In equation (13), the ith block is a prefixed block. However, it is unknown and required to be determined according to the data-dependent selection criteria. The approach is to seek a block which achieved minimized the Euclidean distance $\|w_i(n+1) - w_i(n)\|_2$, thus resulting in the largest magnitude of input sample. Hereby, we can modify the selective block partial-update equation as follows:

$$w(n+1) = w(n) + S(n) \frac{\mu e(n) u}{\|u\|^2 + \delta} + \kappa g(w(n))), \tag{20}$$

where $S(n) = di\ ag(S_1(n), S_2(n), \ldots, S_C(n))$. $S_i(n) E\{0,1\}$, $i = 1, 2 \ldots C$, is a diagonal selection matrix. It can be seen from equation (17) that, at time instant n, the value of $S_i(n)$ will be equal to one and corresponding block $w_i(n)$ in $w(n)$ will be updated while other blocks keep the original value. And when $S(n) = I$, which is an identity matrix, the proposed algorithm will be identical to the BS-NLMS algorithm. The voice activation detection algorithm extracts speech feature parameters at first, and then needs to select specific decision criteria according to the application of the VAD detector to obtain the detection result. Since the third-order statistic of Gaussian noise is zero, for the speech signal whose environmental noise is Gaussian noise, there is a big difference between the third-order statistics of a signal frame and the noise frame, and it is easy to judge the speech segment and the noise segment. Therefore, the third-order statistic can be used to detect the activation of speech. The output result of the VAD algorithm is represented by a Boolean value. When it is judged as a speech frame, the result is "1"; when it is judged as background noise, the result is "0." In the SPU-BS-NLMS algorithm, $w(n)$ is divided into C nonoverlapping blocks which are updated selectively. The elements of $S(n)$ are thus $B = L/C$ equally spaced. Only one block with B tap-weights will be updated per iteration. The computational complexity are summarized in Table 1.

Table 1 shows the computational complexity of BS-NLMS and proposed SPU-BS-NLMS algorithm at each iteration in terms of multiplication, addition, square root, and comparison. The BS-NLMS algorithm requires $2L+6$ multiplications, $4L+2$ additions, $L/P+1$ divisions, L/P comparisons, and L/P square roots for updating full filter coefficients per iteration. And the proposed algorithm requires an extra B comparisons for block selection compared with the BS-NLMS algorithm. However, the saving of $L-B$ multiplications and $2L-2B$ additions will much exceed the additional complexity of B comparision operations for acoustic echo cancellation with large L. The procedure of SPU-PU-NLMS is described in Table 2.

4. Simulations

The unknown system to be identified is of length $L = 800$ and a typical single-clustering of non-zero tap-weights which is randomly generated and located as shown in Figure 2. A zero-mean Gaussian random signal generates an input signal by through a first-order auto regressive (AR) model system $T(z) = 1/(1-0.7z)$. The system superimposes a zero-mean white noise with a Gaussian distribution and a value of 30 dB. The normalized misalignment learning curves are obtained to evaluate the tracking performance of the SPU-BS-NLMS algorithm. The step sizes were chosen as 0.5 for BS-NLMS and 0.6 for SPU-BS-NLMS to get approximately the same normalized mean square deviation after convergence. For the SPU-BS-NLMS algorithm, we set two parameters: $C=2$ and $C=4$.Implement simulation experiments to verify and compare normalized learning curves of BS-NLMS and SPU-BS-NLMS in Figure 3. It is clear from the figure that the SPU-BS-NLMS algorithm with $C=4$ exhibits the slower convergence and the S PU-BS-NLMS algorithm with $C=2$ converges quickly, whereas BS-NLMS obtains the best convergence performance. It can be seen that, when $C=1$, the SPU-BS-NLMS is equal to the BS-NLMS algorithm. When $C=2$ and $C=4$, the algorithm loses some convergence speed with less computational load. The bigger the parmeters of an algorithm, the smaller the computational load and less convergence speed. Thus, the selection of the appropriate C value is mainly based on the trade-off between convergence performance and computational complexity according to specific requirements.The second experiment was conducted to test the proposed algorithm in acoustic echo cancellation in automobiles without double-talk. The experimental platform is the MATLAB environment under Windows. The corpus recording is carried out in an empty room with an area of about 20 m^2. The equipment used for corpus acquisition includes a computer, a microphone, and a speaker. The computer is placed in the middle, and the speaker and the microphone are separated; on both sides and the distance between the two is about 0.45 m. The microphone records the far-end voice signal, which forms an echo through the echo

Table 1: Complexities of the BS-NLMS and SPU-BS-NLMS algorithms.

Algorithm	BS-NLMS	SPU-BS-NLMS
Multiplication	$2L + 6$	$L + B + 6$
Addition	$4L + 2$	$2L + 2B + 2$
Division	$1 + L/P$	$L/P + 1$
Square root	L/P	L/P
Comparison	L/P	$L/P + B$

Table 2: The procedure of SPU-BS-NLMS.

input: $\{x_n, d_n\} = 1, 2, 3, \ldots, L, P, \mu, \alpha, \kappa, \delta$
output: $\{w_n\}_n = 0, 1, 2, \ldots$

Initialization: $w_0 = 0, N = L/P$.
for $n = 0, 1, 2, \ldots$
  $e_n = d_n - x_n^T w_n$
  for $i = 1, 2, \ldots, N$

$$E_i = \left(\sum_{j=(i-1)P+1}^{iP} |w_{j,n}|^2\right)^{1/2}$$

$$S_i = \begin{cases} 1, & (n + j)mo\ dC = 0 \\ 0, & \text{othersize} \end{cases}$$

  end for
  for $j = 1, 2, \ldots, L$
    $g_j(w_n) = 2\alpha^2 w_{j,n} - 2\alpha w_{j.n} \max(1/E_{j/p} + \delta, \alpha)$
    $w_{j,n+1} = w_{j,n} + S_{j,n} u e_n x_{n,j+1} / \|X_{n,y}\|^2 + \delta + k g_j(w_n)$
  end for
end for

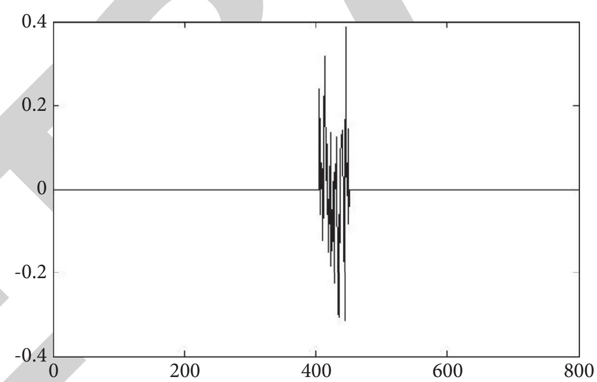


Figure 2: A single cluster block-sparse system tested in experiments.

channel, and is picked up by the microphone together with the near-end speech and ambient noise to become the desired signal. The sampling rate of the data is set at 8000 Hz, and the adaptive FIR filter order is set according to the truncated pulse of the actual room. Two algorithms, the BS-NLMS and the proposed algorithm are compared in simulation. Figure 4(a) calculated the skewness of echo signal and Figure 4(b) depicts echo signal with VAD decision value, which is plotted as dotted line. Additionally, the VAD value is obtained after multimedia smoothing to echo signal. The influence of the echo signal mixed with 10 dB white noise is shown in Figure 5. Figure 5(a) calculated the skewness of echo mixed with 10 dB white noise and Figure 5(b) depicts amplitude of echo mixed with 10 dB white noise and VAD decision value. Figure 6 plots the AEC output employing the SPU-BS-NLMS and the BS-NLMS algorithms. It can be clearly observed that there is only little difference between the two results. The two waveforms with different gray scales represent the two algorithms, respectively. It can be seen from the figure that the residual echo of the proposed SPU-BS-NLMS algorithm is less than that of the conventional algorithm. The residual echo can be observed from the amplitude values of the spectrogram of the time domain.

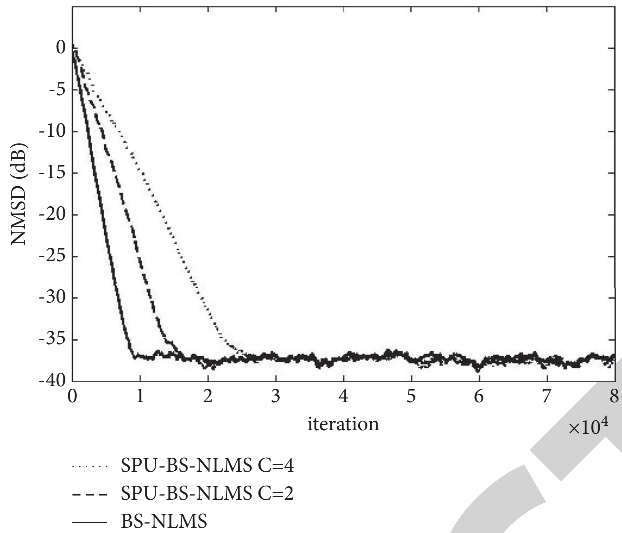
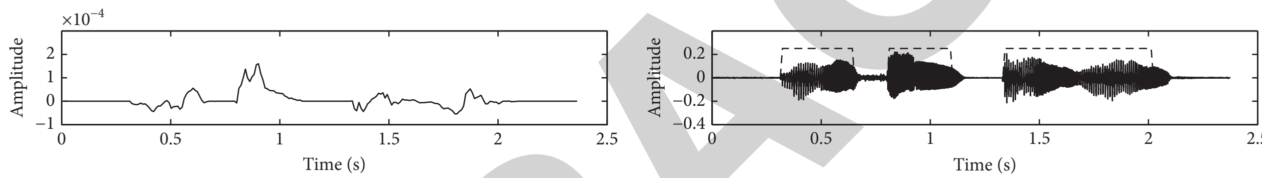
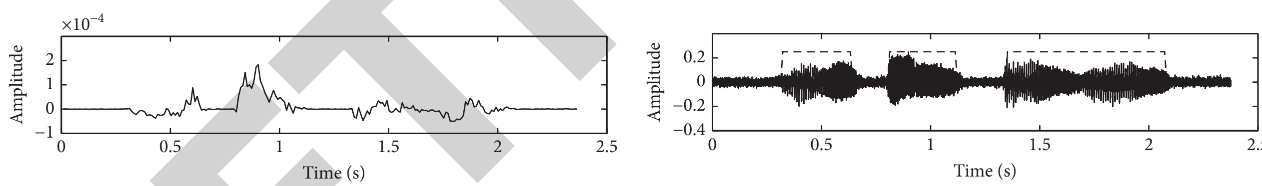


Figure 3: Normalized learning curves of the BS-NLMS and SPU-BS-NLMS with C equal to 2.



(a) (b)

Figure 4: (a) Echo's skewness. (b) Echo's amplitude and VAD decision value.



(a) (b)

Figure 5: (a) Skewness of echo mixed with 10 dB white noise. (b) Amplitude of echo mixed with 10 dB white noise and VAD decision value.

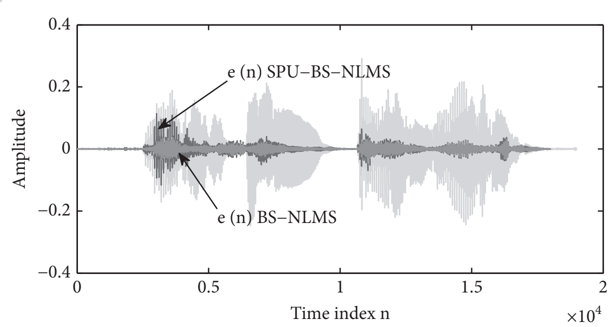


Figure 6: AEC output $e(n)$ SPU-BS-NLMS and BS-NLMS algorithms.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9784859, 1 page
https://doi.org/10.1155/2023/9784859

Retraction

Retracted: Public Sector Performance Assessment Based on DEA Model: The Case of Environmental Policy

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] R. Zhang, "Public Sector Performance Assessment Based on DEA Model: The Case of Environmental Policy," *Security and Communication Networks*, vol. 2022, Article ID 7240379, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7240379, 12 pages
https://doi.org/10.1155/2022/7240379

Research Article

Public Sector Performance Assessment Based on DEA Model: The Case of Environmental Policy

Rongguang Zhang

Business School, Zhengzhou Sias University, Zhengzhou, Henan 450000, China

Correspondence should be addressed to Rongguang Zhang; 15004@sias.edu.cn

Received 28 February 2022; Revised 28 March 2022; Accepted 5 April 2022; Published 28 April 2022

Academic Editor: Chin-Ling Chen

The contradiction between economic development and environmental protection is a core issue that plagues developing countries. The Chinese government has continuously implemented different environmental policies to alleviate the contradiction between economic development and environmental protection, but how they will perform remains unknown. Therefore, this paper revises the classical DEA model and then empirically evaluates the efficiency of environmental policies in 31 Chinese provinces from 2015 to 2020. It is found that the environmental efficiency of each local government has steadily improved, and the provincial differences show a trend of continuous reduction. However, when decomposed, the technical efficiency gap among provinces is large, and the scale efficiency even shows a downward trend. This paper also adopts the Malmquist index for dynamic analysis of environmental efficiency, and the research findings have important theoretical value for scientific formulation of environmental policies.

1. Introduction

Socioeconomic development is accompanied by increasingly serious environmental problems, and some regions are pursuing short-term economic growth at the expense of resources and the environment, causing enormous pressure on the ecological environment, which in turn leads to the deterioration of human living environment. Water pollution, air pollution, land desertification, acid rain, extreme weather, and other environmental problems have emerged, such as the increasingly serious problem of haze, which has caused great inconvenience to people's production and life. The important reason for such problems is that the previous sloppy economic growth mode of pursuing economic development unilaterally and the wrong development concept of polluting first and treating later have caused irreversible process of environmental destruction. Environmental problems have become a major obstacle to the sustainable development of China's economy and society. The government is the main body of environmental governance, and environmental governance is an important part of national governance; along with the increasing severity of environmental problems, governments at all levels have begun to pay attention and focus on how to achieve a balance between economic development and environmental protection [1]. The report of the 18th National Congress of China has elevated the construction of ecological civilization to a national macrostrategy, and the report of the 19th National Congress emphasizes that "green water and green mountains are the silver mountain of gold." At present, China has proposed a carbon neutral strategy, striving to reach the peak of carbon dioxide emissions by 2030 and achieving carbon neutrality by 2060. Various local governments have put environmental governance in the forefront, but how effective it will be is not yet known, and a scientific and systematic study is urgently needed. More importantly, environmental governance is the primary task of the governmental public sector and an important part of evaluating the performance of the governmental public sector.

Scientific performance evaluation methods are a key part of implementing government performance evaluation of low-carbon governance. Scholars emphasize that government performance evaluation should choose different methods according to different evaluation purposes and result users [2, 3]. Among them, the DEA method shows unique advantages in evaluating the efficiency of public administrations and organizations and is widely used by scholars in the evaluation of government governance performance and energy and environmental efficiency [4–6], and scholars believe that the DEA method is an effective method that can combine information on the main services (outputs) and inputs provided by government service providers for measurement. However, the classical DEA model relies too much on the chosen indicators for efficiency evaluation and is slightly inadequate in dynamic evaluation capability. Therefore, this paper constructs a local government low-carbon governance performance evaluation index system from the perspective of local government low-carbon governance performance, using a combination of nonradial DEA method, RAM model and Malmquist index, and evaluates local government low-carbon governance performance statically and dynamically, with a view to providing the government with environmental governance formulation. It is intended to provide theoretical references for the government to make environmental governance decisions.

2. Methodology

In order to integrate energy, carbon emissions, and economic growth into the framework of efficiency analysis, we first need to construct a production possibility set that contains both desired and undesired outputs [7, 8]. In the literature, normal products are often defined as desired outputs, and pollutants emitted as byproducts of industrial production, such as waste gas and waste water, are defined as undesired outputs [9, 10]. Scholars refer to the technological structure of the relationship between outputs, including pollutant byproducts, and factor resource inputs as environmental technology [11]. Environmental technology is broken down by output, but the impact of input breakdown on the technology structure relationship is not considered. As the material basis for pollutant production, energy does not fit directly into the construct of best practice facets as do other resources [12]. Therefore, we further disaggregate production inputs into energy elements and common elements such as capital, equipment, and labor. Assume that each region uses N common inputs $x = (x1, \ldots, xN) \in R_N^+$ and M energy inputs $e = (e1, \ldots, eN) \in R_M^+$, which gets P desired outputs $y = (y1, \ldots, yN) \in R_P^+$ and I undesired outputs $b = (b1, \ldots, bN) \in R_I^+$. The production possibility set is simulated as

$$T = \left\{(x, e, y, b):\ (x, e) \text{ can produce } (\mathrm{y}, \mathrm{b}), x \in R_+^N, e \in R_+^M\right\}. \tag{1}$$

Assuming that the input and output vectors for each province are $J\,(xtj, etj, ytj, btj)$ at each period $t = 1, \ldots, T$, and $j = 1, \ldots, J$, we use the DEA extension model-RAM model to construct the optimal practice frontier (or reference technology) for China and compare the production of each province with the optimal practice frontier to measure the change in efficiency.

2.1. Efficiency Model Based on Desired Output. Scholars were the first to produce optimal practice bounds by constructing a nonparametric linear convex surface [13]. Since then, DEA methods represented by the CRS model based on constant payoffs to scale and the VRS model with variable payoffs to scale have followed the definition of technical efficiency, measuring the ability to achieve maximum output given various input factors or minimize inputs for a given level of output [14, 15]. The horizontal and vertical axes in Figure 1 are represented as two input factors $x1$ and $x2$, respectively, and SS represents the optimal production possible frontier that can be achieved without efficiency losses. The technical efficiency of decision unit j is equal to OJ'/OJ. In fact, the inefficiency of decision unit j consists of a combination of excess JJ' of input factors due to technical inefficiency and slack variables CJ' due to improper input factors, i.e., input factor X1 while maintaining the same output can continue to decrease from point J' to point C. The true efficiency of decision unit A should be OJ'/(JC + OJ'). Therefore, the traditional DEA ignores the interfactor allocation efficiency due to the radial limitation.

The RAM model characterizes TE based on the slackness of inputs and outputs with respect to the projection of the efficiency frontier and allows for free variation in the amount of input and output factors, instead of requiring the input and output factors to vary in the same proportion as in the traditional DEA solution [16, 17]. At the same time, the RAM model does not require the choice of calculating efficiency values from the input perspective or from the output perspective. More importantly, each input-output variable is summed in the objective function of the RAM planning model in a structure that allows for independent efficiency measures of desired or undesired outputs as well as integration of desired and undesired output efficiencies in the same model architecture. In addition, since the RAM constraints are equated, the slack in each element represents the difference between the current use of that element and the use in the optimal state of technology, facilitating the resolution of the sources of inefficiency. Assume that the jth province in period $t\,(j = 1, \ldots, J)$ has input and normal output slack relative to the production frontier projection of $S_n^x \geq 0, \forall_n; S_p^y \geq 0, \forall_p$. Thus, the efficiency model based on undesired output is [18]

$$\max\left\{\left(\sum_{n=1}^{N} R_n^x s_n^x + \sum_{p=1}^{P} R_p^y s_p^y\right) \Bigg| \begin{array}{l} \sum_{j=1}^{J} x_{nj}\lambda_j + s_n^x = x_{nj}, \forall n; \sum_{j=1}^{J} y_{pj}\lambda_j - s_p^y = y_{pj}, \forall p; \\ \sum_{j=1}^{J} \lambda_j = 1, \lambda \geq 0, \forall j; s_n^x \geq 0, \forall n; s_p^y \geq 0, \forall p; \end{array}\right\}. \tag{2}$$

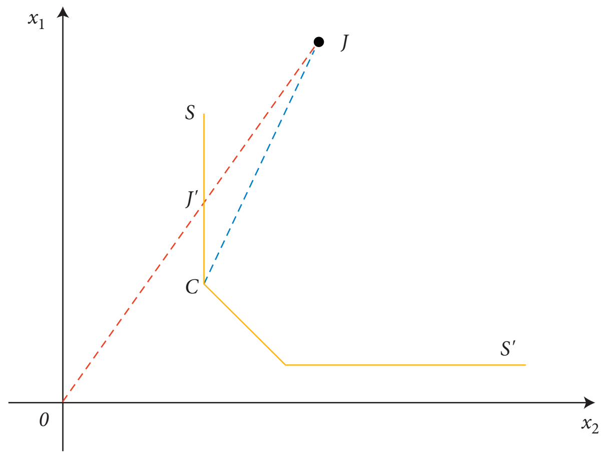


Figure 1: Traditional DEA model.

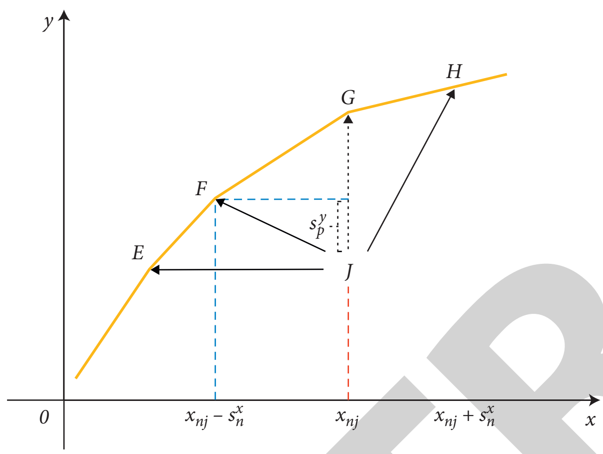


Figure 2: RAM efficiency (1).

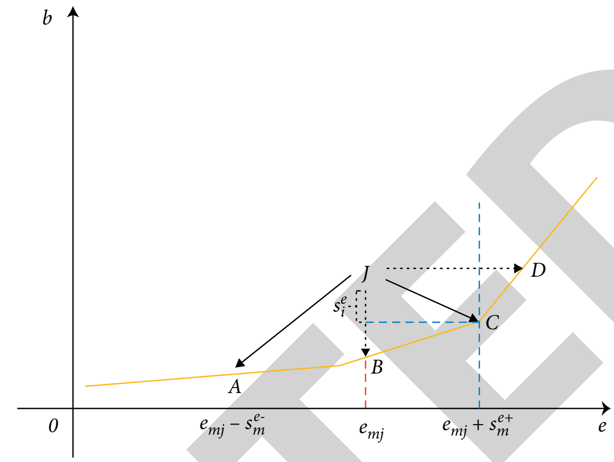


Figure 3: RAM efficiency (2).

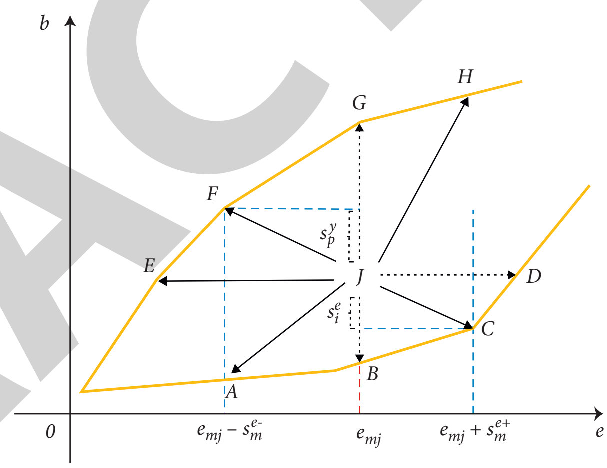


Figure 4: RAM joint efficiency.

The adjustment interval for slack (S_n^x, S_p^y) is defined based on the calculation of the extreme differences [max(xnj)-min(xnj)] and [max(ypj)-min(ypi)] for each input-output in all evaluated provinces.

$$R_n^x = \frac{1}{(N+P)\left[Max\left(x_{nj}\right) - Min\left(x_{nj}\right)\right]}. \tag{3}$$

$$R_p^y = \frac{1}{(N+P)\left[Max\left(y_{pj}\right) - Min\left(y_{pj}\right)\right]}. \tag{4}$$

From the definition of relaxation it follows that (S_n^x, S_p^y) lies between zero and the extreme difference.

$$0 \le s_n^{x*} = x_{nj} - \sum_{j=1}^{J} x_{nj}\lambda^* \le R_n^x. \tag{5}$$

$$0 \le s_p^{y*} = \sum_{j=1}^{J} y_{pj}\lambda^* - y_{pj} \le R_p^y, \tag{6}$$

where $*$ denotes the state in which the model achieves an optimal solution. x is the weight of the cross-sectional observation of the maximum relative efficiency possible in reality for each province when the model achieves an optimal solution. Let $\sum_{j=1}^{J} \lambda_j^t$, combined with the constraint that the weight variable λ_j^t is nonnegative, express the production technology as a variable payoff of scale, then the objective function value of linear programming to maximize the degree of inefficiency satisfies max(-) $\in$ [0, 1], and the RAM economic efficiency indicator for province j in period t can be transformed into

$$0 \le \theta_p = 1 - \left(\sum_{n=1}^{N} R_n^x s_n^{x*} + \sum_{p=1}^{P} R_p^y s_p^{y*}\right) \le 1. \tag{7}$$

$\theta_p \in [0, 1]$ satisfies the boundedness of efficiency values with monotonic orderability. When all input slack and output slack are equal to zero, the objective function value is equal to zero, at which point $\theta_p = 1$, indicating that the region is located on the optimal practice boundary and reaches the technically efficient Pareto optimum. The principle of the economic efficiency model can be explained in Figure 2. In Figure 2, the horizontal axis denotes the input x, the total draw denotes the desired output y, and the arc EFG constitutes the optimal practice boundary of economic efficiency. For a province j, the current input-output point J of that province in the solution process needs to be projected along the JF direction, when economic output y tends to increase while input x tends to

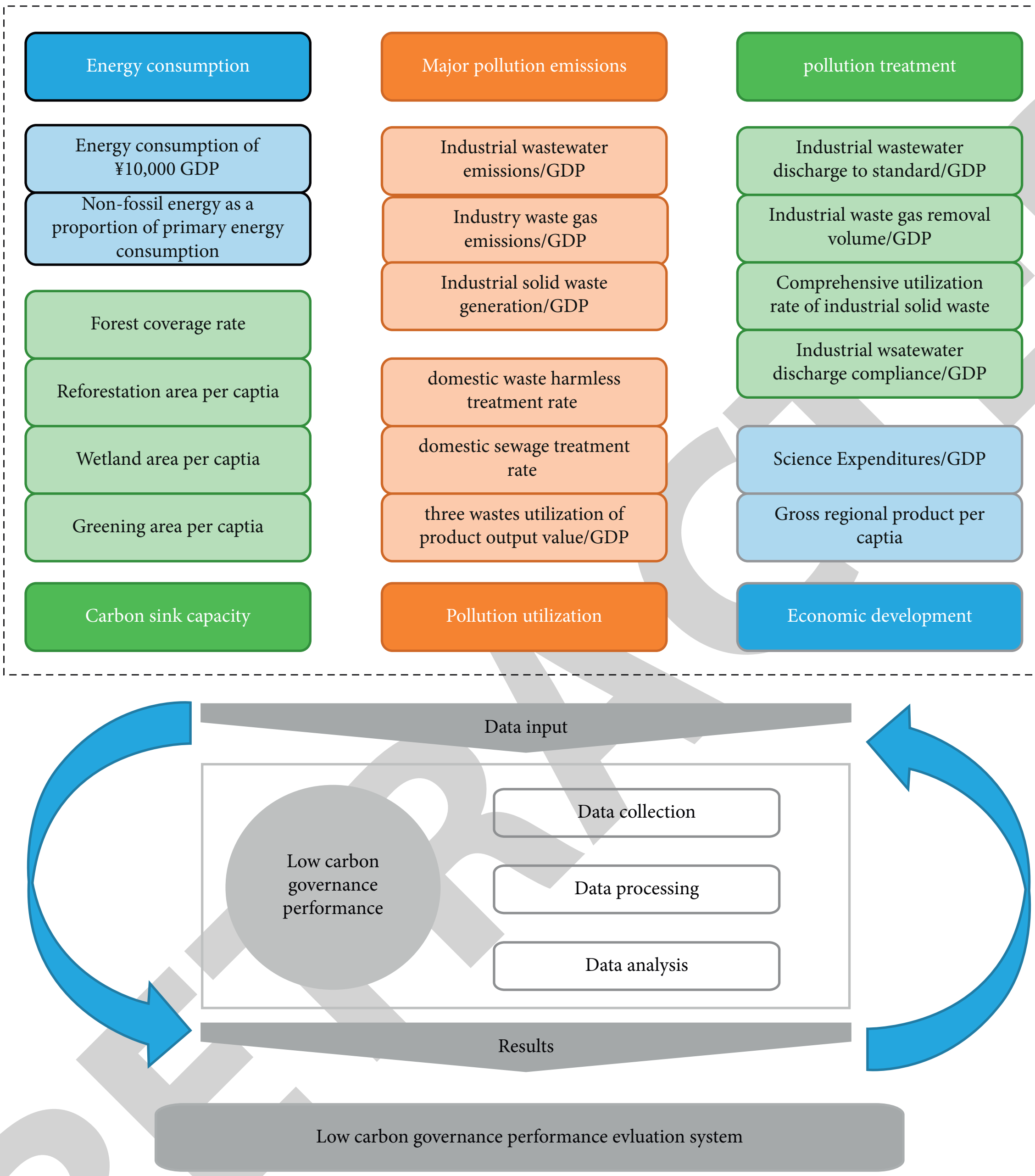


Figure 5: Local government low-carbon governance performance evaluation system.

decrease, and thus efficiency improves. The JH direction is an inefficient projection, when economic output y tends to increase, but the output increase is due to the increase in input factor x, not due to efficiency improvement.

2.2. Environmental Efficiency Model Based on Nondesired Outputs. Further environmental factors are incorporated into the model. The slack variable S_i^b is to represent the excess emissions of the ith pollutant. Combined with the negative sign constraint in front of the slack variables, this satisfies the inverse expectation that the production process tends to expand economic output and reduce environmental pollution. Traditional DEA treats energy as a common input due to perspective constraints [19–21]. The expansion of energy consumption is judged as a deterioration of efficiency, which leads to the fact that energy efficiency does not reflect the productivity impact of intersubstitution between energy factors and other factors or between different energy factors [22]. Scholars using nonradial DEA measures of energy efficiency have found a mixed effect, i.e., an increase in alternative energy use may lead to a contraction of a larger number of inputs of other types of energy, holding other inputs and outputs constant, and an expansion of a certain type of energy consumption does not necessarily indicate a deterioration in efficiency but may instead imply an improvement in efficiency. We construct energy slack variables S_m^{e-} and S_m^{e+} and combine the positive and negative sign constraints in front of the slack variables in the equation constraint to make them indicate two projection directions

Table 1: Low-carbon governance efficiency in China from 2015 to 2017.

Region	2015			2016			2017		
	TE	PTE	SE	TE	PTE	SE	TE	PTE	SE
Jiangsu	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Beijing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Shanghai	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Anhui	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Guangdong	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Guizhou	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Xizang	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Tianjin	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hebei	0.967	1.000	0.967	0.993	1.000	0.993	0.951	1.000	0.951
Shanxi	0.968	0.977	0.991	1.000	1.000	1.000	1.000	1.000	1.000
Neimenggu	0.880	0.902	0.975	0.861	0.872	0.987	0.873	0.875	0.998
Liaoning	0.752	0.816	0.921	0.722	0.808	0.894	0.884	0.900	0.981
Jilin	0.732	0.735	0.996	0.779	0.779	1.000	0.714	0.751	0.950
Heilongjiang	0.716	0.746	0.959	0.710	0.751	0.945	0.735	0.766	0.959
Zhejiang	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Fujian	0.936	1.000	0.936	0.943	1.000	0.943	0.935	1.000	0.935
Jiangxi	0.879	0.884	0.994	0.911	0.911	1.000	0.900	0.913	0.987
Shandong	0.911	1.000	0.911	0.937	1.000	0.937	1.000	1.000	1.000
Henan	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hubei	0.813	0.823	0.987	0.858	0.861	0.996	0.819	0.855	0.958
Hunan	0.814	0.815	1.000	0.927	0.929	0.998	0.935	0.937	0.998
Guangxi	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hainan	0.738	0.739	0.998	0.827	0.835	0.991	0.765	0.799	0.958
Chongqing	0.718	0.759	0.945	0.763	0.874	0.872	0.762	0.891	0.855
Sichuan	0.940	0.941	0.999	0.888	0.894	0.993	0.937	0.937	1.000
Yunnan	0.855	0.864	0.990	0.939	1.000	0.939	0.984	1.000	0.984
Shaanxi	0.842	0.880	0.957	0.825	0.872	0.945	0.847	0.888	0.953
Gansu	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Qinghai	0.901	0.986	0.910	0.944	1.000	0.944	0.937	0.986	0.951
Ningxia	0.714	0.757	0.940	0.762	0.795	0.958	0.783	0.807	0.971
Xinjiang	0.911	0.986	0.924	0.888	1.000	0.888	0.969	0.970	1.000
Mean	0.903	0.923	0.978	0.919	0.941	0.975	0.927	0.944	0.980

of energy expansion and contraction; i.e., we define the environmental efficiency model for nondesired outputs as

$$\max\left\{\begin{array}{c}\left(\sum_{n=1}^{N} R_n^x s_n^x+\sum_{m=1}^{M} R_m^e\left(s_m^{e+}+s_m^{e-}\right)+\sum_{i=1}^{I} R_i^b s_i^b\right) \mid \sum_{j=1}^{J} x_{nj}\lambda_j \\ +s_n^x=x_{nj}, \forall n; \sum_{j=1}^{J} e_{mj}\lambda_j-s_m^{e+}+s_m^{e-}=e_{mj}, \forall m; \\ \sum_{j=1}^{J} b_{ij}\lambda j+s_i^b=b_{ij}, \forall i; \sum_{j=1}^{J}\lambda_j=1, \lambda_j\geq 0, \forall j; \\ s_n^x\geq 0, \forall n; s_m^{e+}\geq 0, s_m^{e-}\geq 0, \forall m; s_i^b\geq 0, \forall i\end{array}\right\}. \tag{8}$$

When limiting the nondesired output to carbon emissions an environmental emitter, the RAM carbon environmental efficiency (CE) indicator for province j in period t is obtained as follows:

$$0\leq\theta_E=1-\left(\sum_{n=1}^{N} R_n^x s_n^{x*}+\sum_{m=1}^{M} R_m^e\left(s_m^{e+*}+s_m^{e-*}\right)+\sum_{i=1}^{I} R_i^b s_i^{b*}\right)\leq 1. \tag{9}$$

Figure 3 represents the environmental efficiency model. Assuming that input factor energy is applicable, the horizontal axis represents energy e and the vertical axis represents nondesired output b. For a province j, the point J corresponds to carbon emission cj and energy consumption emj. Since the energy mixing effect is considered, the RAM carbon environmental efficiency has two effective projection directions JA and JC, and accordingly the optimal practice boundary of environmental efficiency extends from the arc BC to CD. During the projection along the JA direction to the optimal boundary, both energy consumption and carbon emission tend to decrease. In the projection along JC direction towards the optimal boundary, the consumption of certain energy sources tends to increase but causes a decrease in carbon emissions due to the presence of the mixing effect.

2.3. Joint Efficiency Model Based on Dual Output. Economic efficiency (PE) assumes that there is no environmental control and pursues economic efficiency while ignoring environmental pollution, while carbon environmental efficiency (CE) meets the reverse expectation that the production process tends to reduce environmental pollution through the improvement of energy use and the optimal allocation of other factors of production, and its efficiency connotes the implementation of energy-saving and carbon-reducing environmental controls. The use of PE or CE alone

Table 2: Low-carbon governance efficiency in China from 2018 to 2020.

Region	2018			2019			2020		
	TE	PTE	SE	TE	PTE	SE	TE	PTE	SE
Jiangsu	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0000	1.000
Beijing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0000	1.000
Shanghai	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Anhui	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Guangdong	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Guizhou	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Xizang	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Tianjin	1.000	1.000	1.000	0.938	1.000	0.938	0.942	1.000	0.942
Hebei	0.791	0.852	0.928	0.941	1.000	0.941	0.912	0.913	0.999
Shanxi	0.777	0.989	0.786	1.000	1.000	1.000	0.978	1.000	0.978
Neimenggu	0.764	0.858	0.891	0.925	1.000	0.925	0.810	0.856	0.947
Liaoning	0.912	0.912	1.000	0.875	0.884	0.990	1.000	1.000	1.000
Jilin	0.594	0.751	0.791	0.755	0.875	0.862	0.690	0.725	0.952
Heilongjiang	0.590	0.771	0.765	0.773	0.774	0.999	0.739	0.777	0.951
Zhejiang	1.000	1.000	1.000	0.988	1.000	0.988	0.960	1.000	0.960
Fujian	1.000	1.000	1.000	0.904	1.000	0.904	1.000	1.000	1.000
Jiangxi	0.804	0.915	0.878	0.870	0.882	0.987	0.831	0.877	0.947
Shandong	0.976	1.000	0.976	0.948	1.000	0.948	1.000	1.000	1.000
Henan	0.933	0.968	0.964	1.000	1.000	1.000	0.900	0.901	0.999
Hubei	0.680	0.817	0.832	0.799	0.800	0.999	0.859	0.861	0.998
Hunan	0.748	0.870	0.859	0.861	0.890	0.968	0.834	0.844	0.988
Guangxi	0.998	1.000	0.998	0.993	1.000	0.993	1.000	1.000	1.000
Hainan	0.622	0.809	0.769	0.741	0.773	0.959	0.790	0.812	0.973
Chongqing	0.628	0.845	0.743	0.715	0.738	0.968	0.716	0.771	0.929
Sichuan	0.722	0.918	0.787	0.897	0.904	0.993	0.821	0.884	0.929
Yunnan	0.783	0.948	0.825	0.908	0.912	0.995	0.897	0.897	0.999
Shaanxi	0.755	0.915	0.825	0.895	0.898	0.996	0.810	0.913	0.888
Gansu	0.949	1.000	0.949	1.000	1.000	1.000	1.000	1.000	1.000
Qinghai	0.866	0.936	0.925	0.872	0.879	0.992	0.758	0.829	0.914
Ningxia	0.672	0.799	0.841	0.805	0.816	0.987	0.794	0.806	0.986
Xinjiang	0.846	0.951	0.890	0.908	1.000	0.908	0.882	0.888	0.994
Mean	0.852	0.930	0.910	0.913	0.936	0.975	0.901	0.921	0.977

as indicators of economic development is clearly one-sided, so it is necessary to integrate them in a unified framework. The joint efficiency model is as follows:

$$\max\left\{\begin{array}{c}\left(\sum_{n=1}^{N}R_n^x s_n^x+\sum_{m=1}^{M}R_m^e\left(s_m^{e+}-s_m^{e-}\right)+\sum_{p=1}^{P}R_p^y s_p^y+\sum_{i=1}^{I}R_i^b s_i^b\right)\mid\\ \sum_{j=1}^{J}x_{nj}\lambda_j+s_n^x=x_{nj},\forall n;\\ \sum_{j=1}^{J}e_{mj}\lambda_j-s_m^{e+}+s_m^{e-}=e_{mj},\forall m;\ \sum_{j=1}^{J}y_{pj}\lambda_j-s_p^y=y_{pj},\forall p;\\ \sum_{j=1}^{J}b_{ij}\lambda_j+s_i^b=b_{ij},\forall i;\\ \sum_{j=1}^{J}\lambda_j=1,\ \lambda_j\geq 0,\forall j;\ s_n^x\geq 0,\forall n;\ s_m^{e+}\geq 0,\ s_m^{e-}\geq 0,\forall m;\\ s_p^y\geq 0,\forall p;\ s_i^b\geq 0,\forall i\end{array}\right\}. \tag{10}$$

The planning model of equation (10) considers both desired and undesired outputs, and accordingly, two optimal practice bounds can be constructed, one for carbon emissions and the other for economic growth. The joint efficiency can measure the degree of coupling between economic growth and energy-saving and carbon-reduction control, which reflects the low-carbon economy model of

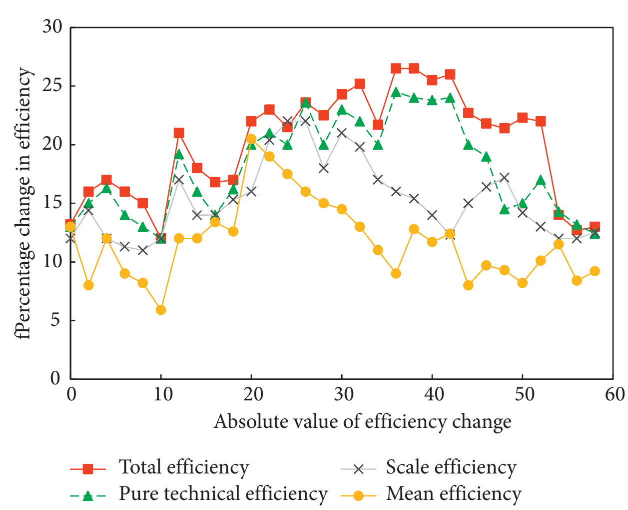


Figure 6: Trend change in efficiency distribution (TE, PTE, SE).

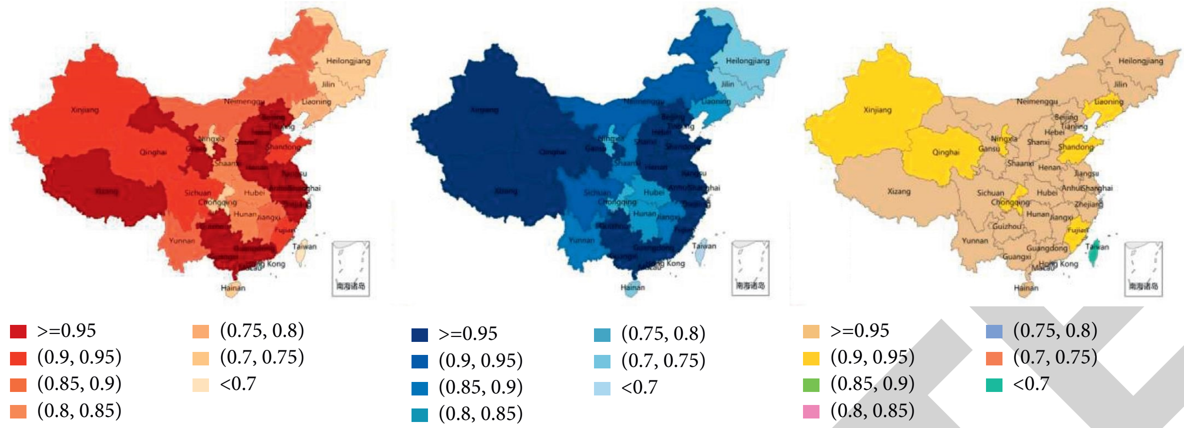


Figure 7: 2015 low-carbon governance efficiency in China (TE, PTE, SE).

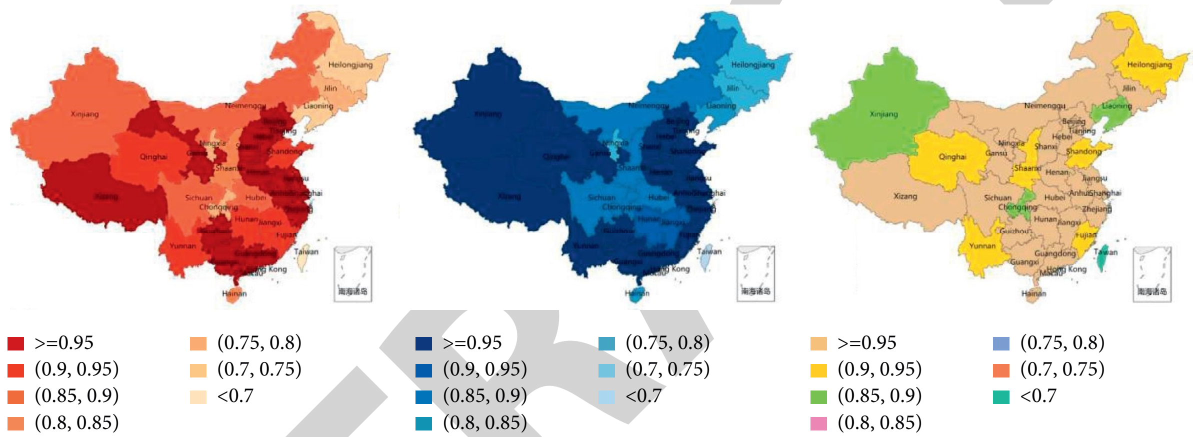


Figure 8: 2016 low-carbon governance efficiency in China (TE, PTE, SE).

win-win situation between economic and social development and ecological protection. At this point, the joint efficiency index of province j at time t can be calculated by the following equation:

$$0 \le \theta_U = 1 - \left(\sum_{n=1}^{N} R_n^x s_n^{x*} + \sum_{m=1}^{M} R_m^e \left(s_m^{e+*} + s_m^{e-*} \right) + \sum_{p=1}^{P} R_p^y s_p^{y*} + \sum_{i=1}^{I} R_i^b s_i^{b*} \right) \le 1. \tag{11}$$

This is because the existence of two optimal practice bounds that the efficiency projection of the decision unit in the joint efficiency model changes from the unilateral projection under the economic efficiency model and the environmental efficiency model to the bilateral projection. In Figure 4, the horizontal axis represents energy e and the vertical axis represents both normal output y and carbon emissions c. The arc ABCD represents the optimal practice boundary for carbon emissions and the arc EFG represents the optimal practice boundary for economic growth. The low-carbon economy model cannot unilaterally pursue economic growth at the expense of the environment, nor can it achieve carbon reduction at the expense of economic growth. Specifically, province J can improve economic efficiency along the JF direction, when normal output increases from *ypj* to *ypj*+ S_i^b while the energy consumption level decreases from *emj* to *emj*- S_m^{e-}. Second, the province J can improve the carbon environmental efficiency along the JC and JA projection directions. jC direction implies management improvement, which improves the energy use efficiency by optimizing the energy structure or applying low-carbon technology equipment, when the energy consumption level increases from *emj* to *emj*+ S_m^{e+}, but the carbon emission is reduced. ja direction is a natural emission reduction, which any region can achieve without management improvement by limiting. The JA direction is a natural reduction in energy consumption and carbon emissions that can be achieved in any region.

In sum, the revised DEA model can provide more accurate information and can also better predict the efficiency of low-carbon environmental governance policies.

2.4. Malmquist Index. Malmquist index is able to measure the dynamic efficiency of the environmental policy, which makes up for the shortage of traditional DEA models that can only analyze static efficiency. (x^t, y^t) denotes the input and output quantities in period t, $D_c^t(x^t, y^t)$ denotes the

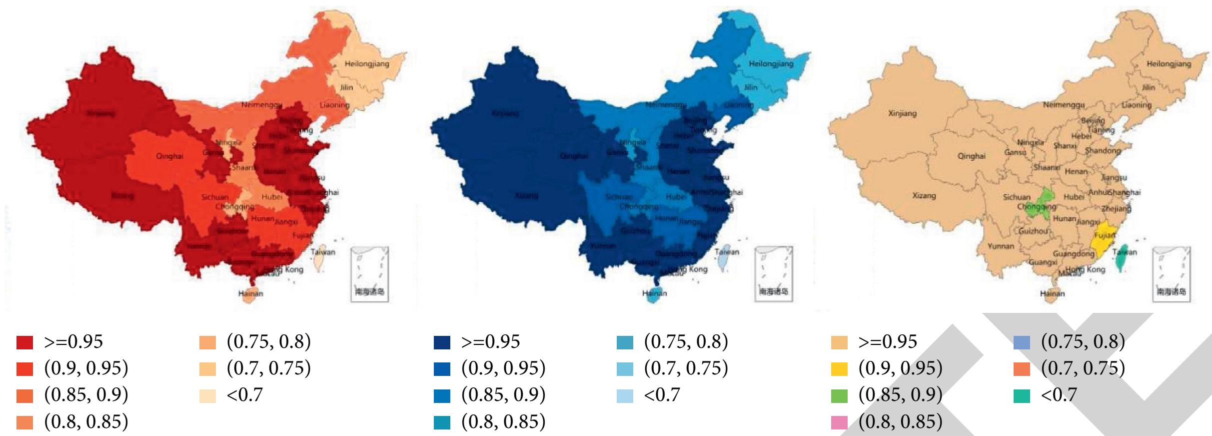


Figure 9: 2017 low-carbon governance efficiency in China (TE, PTE, SE).

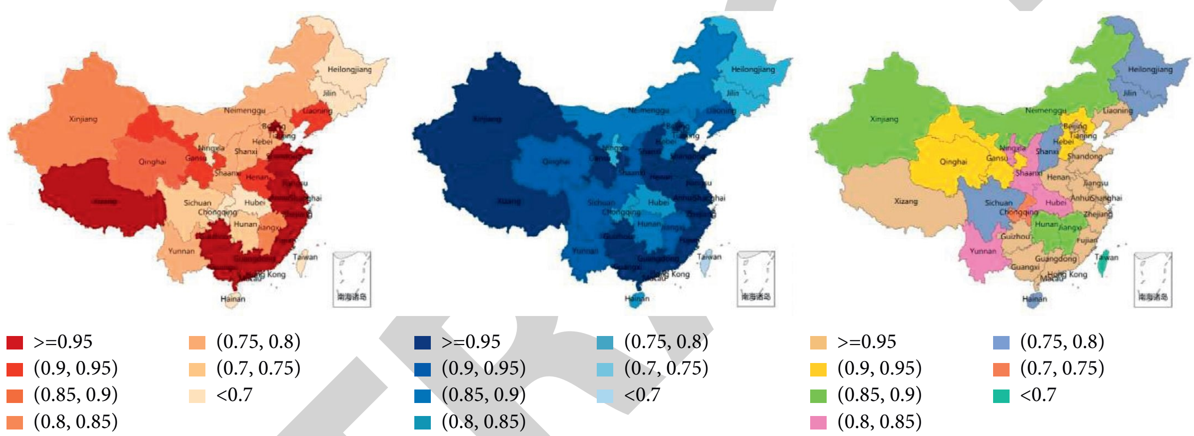


Figure 10: 2018 low-carbon governance efficiency in China (TE, PTE, SE).

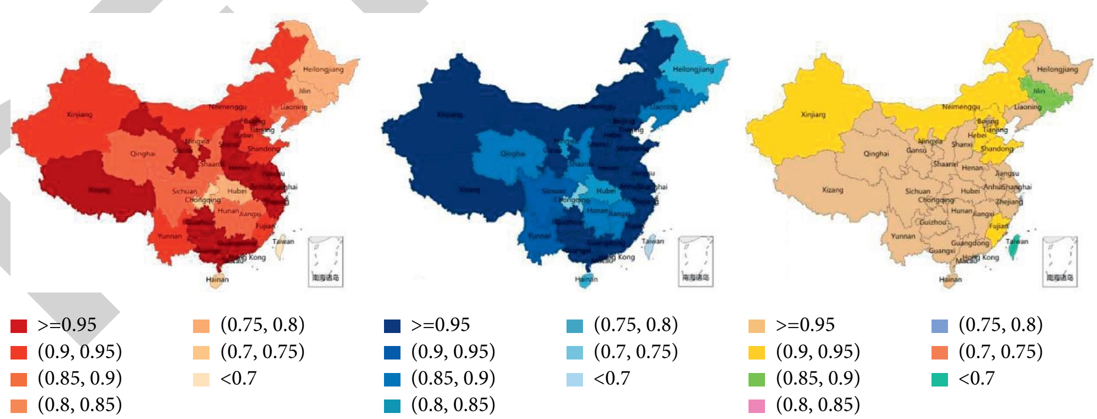


Figure 11: 2019 low-carbon governance efficiency in China (TE, PTE, SE).

output distance function under the technical conditions in period t, M^t denotes the value of change in efficiency from period t to period $t+1$ under the technical conditions in period t, and (x^{t+1}, y^{t+1}) denotes the input and output quantities in period $t+1$, $D_c^{t+1}(x^{t+1}, y^{t+1})$ denotes the output distance function under the technical conditions in period $t+1$, and M^{t+1} denotes the value of change in efficiency from period t to period $t+1$ under the technical conditions in period $t+1$. The Malmquist index formula is as follows [23]:

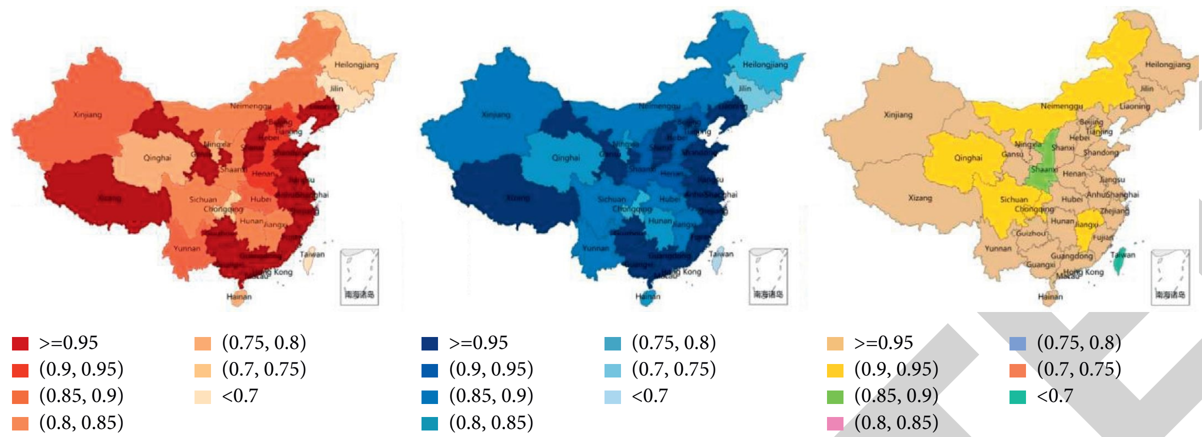


Figure 12: 2020 low-carbon governance efficiency in China (TE, PTE, SE).

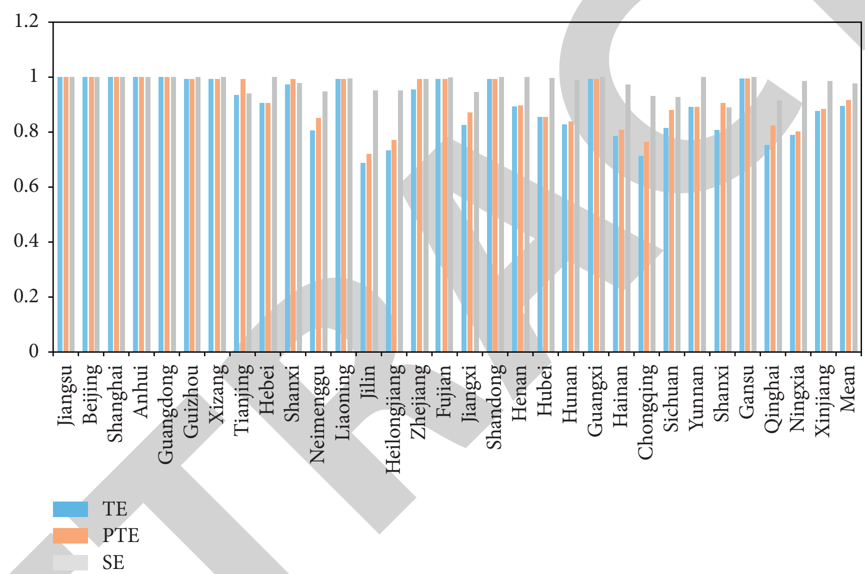


Figure 13: The efficiency values from in 2020.

$$
\begin{aligned}
TFP &= M^{t+1}\left(x^{t+1}, y^{t+1}, x^{t}, y^{t}\right)\\
&= \left(M^{t} \times M^{t+1}\right)^{\frac{1}{2}}\\
&= \left[\frac{D_c^{t}\left(x^{t+1}, y^{t+1}\right)}{D_c^{t}\left(x^{t}, y^{t}\right)} \times \frac{D_c^{t+1}\left(x^{t+1}, y^{t+1}\right)}{D_c^{t+1}\left(x^{t}, y^{t}\right)}\right]^{\frac{1}{2}}.
\end{aligned}
\tag{12}
$$

When TFP >1, it indicates an upward trend; when $TFP = 1$, it indicates no change; and when TFP <1, it indicates a downward trend. Assuming constant returns to scale, TFP can be further decomposed into technical efficiency change ($EFFCH$) and technical progress change ($TECH$). Assuming variable returns to scale, EFFCH can be decomposed again into pure technical efficiency change ($PECH$) and scale efficiency change ($SECH$). Therefore, the TFP is calculated as

Table 3: The Malmquist index and its decomposition from 2015 to 2020.

Period	EFFCH	TECH	PECH	SECH	TFP
2015–2016	1.0024	0.9328	1.0029	0.9996	0.9349
2016–2017	1.0104	0.9061	1.0038	1.0063	0.9152
2017–2018	0.9140	1.0957	0.9849	0.9277	0.9955
2018–2019	1.1111	0.8570	1.0226	1.0863	0.9438
2019–2020	0.9682	1.0459	0.9708	0.9976	1.0109
Mean	1.0012	0.9675	0.9970	1.0035	0.9601

$$
TFP = EFFCH \times \mathrm{TECH} = PECH \times SECH \times \mathrm{TECH}. \tag{13}
$$

2.5. Index System Construction. Based on the extensive literature review, we selected energy consumption, major pollution emissions, pollution control and utilization, carbon sink capacity, and economic development as our index system. Figure 5 shows the index system this paper used, energy consumption. It is the goal of local governments to

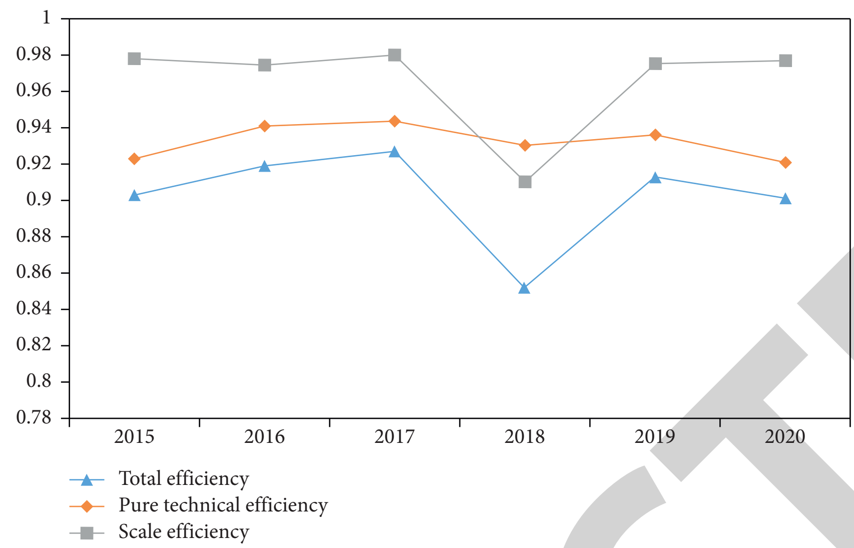


FIGURE 14: The mean efficiency values from 2015 to 2020.

achieve more economic output with less energy consumption by formulating relevant policies to guide, control, and regulate the behavior of social agents such as enterprises and the public, and it is used to reflect the government's function of promoting energy efficiency. Therefore, two main indicators are chosen for this dimension: energy consumption per 10,000 Yuan of GDP (equivalent value) and the share of nonfossil energy in primary energy consumption [24].

Major pollution emissions: The rapid development of China's economy over the past three decades is mainly attributed to the energy- and labor-intensive industrial sector, which has not only increased the emissions of various greenhouse gases such as carbon dioxide, but also led to serious environmental pollution. The industrial sector will inevitably play an important role in promoting a low-carbon, resource-efficient, and environment-friendly society in China. The environmental performance of industry is considered to be an important criterion for evaluating the environmental performance of local governments. Therefore, this indicator layer mainly selects industrial "three waste" emission indicators, including industrial wastewater emission, industrial gas emission, and industrial solid waste generation [25].

Pollution control and utilization: This is the ability of local governments to remove and utilize pollutants by investing certain human and financial resources to guide, control, and regulate the behavior of enterprises and the public and other social entities. The indicators of industrial "three wastes" emission in (2) correspond to the indicators of industrial "three wastes" pollution control and utilization adopted in this part, mainly including industrial wastewater discharge compliance, industrial waste gas removal, industrial solid waste comprehensive utilization rate, "three wastes" comprehensive utilization products output value, and "three wastes" comprehensive utilization products output value, with (4) carbon sink capacity, the value of the "three wastes" comprehensive utilization products, the domestic sewage treatment rate, the domestic waste harmless treatment rate [26].

Carbon sink capacity: This is the ability of local governments to store atmospheric greenhouse gases in biological carbon pools through land use adjustment and forestry measures. "Carbon sink" is the deposit of "carbon" in nature, and the largest deposit of "carbon" on Earth is forest vegetation. Therefore, in this paper, the indicators of green area per capita in built-up areas, wetland area per capita, afforestation area per capita in the current year, green coverage rate in built-up areas, and forest coverage rate are chosen to measure this function of local governments to enhance carbon sink capacity [27].

Economic development: The ultimate goal of low-carbon governance is to achieve sustainable economic development, and since technological progress is one of the determinants or controlling factors of a low-carbon economy, carbon productivity is also determined by the level of technology. Therefore, drawing on the research results of previous scholars, this paper selects the gross regional production value per capita, scientific expenditure, and the number of employees in science and technology services to measure the ability of local governments to promote economic development [28].

3. Results and Discussion

As shown in Tables 1 and 2, the total efficiency (TE) of low-carbon governance in Jiangsu, background, injury, Anhui, Guangdong, Guizhou, and Tibet is basically 1, reaching DEA effective. Tianjin, Zhejiang, Shandong, Henan, Guangxi, and Gansu are slightly less efficient. Jilin and Heilongjiang have lower total efficiency, with an average value below 0.72 and an annual minimum value less than 0.60. Figure 6 illustrates the trends in total efficiency, pure technical efficiency, and scale efficiency.

Overall, the provinces do not differ much and the gap tends to narrow gradually. In terms of pure technical efficiency (PTE), Jiangsu, Beijing, Tianjin, Shanghai, Zhejiang, Anhui, Fujian, Shandong, Guangdong, Guangxi, Guizhou, Tibet, and Gansu all reached DEA effective. Hebei, Shanxi, and Henan out of technical efficiency are slightly inferior. The national pure base efficiency reached 0.933, indicating that the provincial governments in China operate at a high level of efficiency in low-carbon governance. In terms of scale efficiency (SE), Jiangsu, Beijing, Injury, Anhui, Guangdong, Guizhou, and Tibet all reach 1 per year, and the average value of scale efficiency of 23 provinces is above 0.95, indicating that the scale of low-carbon governance inputs in China has reached the optimal scale (Figures 7-12).

As shown in Figure 13, the total efficiency in Jilin and Chongqing is the lowest across the whole provinces. Jiangsu, Beijing, Shanghai, Anhui, and Guangdong have the highest efficiency getting the effective DEA.

Based on the measurement and analysis of government low-carbon governance TE, the evolutionary path of government low-carbon governance efficiency over time is further examined by Malmquist index and its decomposition, and the measurement results are shown in Table 3. The mean value of TFP is 0.9601, which indicates that government low-carbon governance performance is at a slight decline level. From the dynamic decomposition results, both EFFCH and PECH are greater than 1, indicating that technical efficiency change and pure technical efficiency change are the main reasons for the government's low-carbon governance efficiency improvement (Figure 14).

4. Conclusions

Environmental issues have become a hot issue in the field of national governance in China and gradually received the attention and importance of governments at all levels. China has mentioned the construction of ecological civilization in the report of the 18th National Congress and the report of the 19th National Congress, and ecological civilization has become a national strategic choice. However, the success of environmental governance work depends on not only the quantity and strength of environmental governance resources invested, but more importantly, the effectiveness of the use of environmental governance resources. In other words, the issue of environmental governance has become the most important aspect in evaluating government efficiency. So, what is the low-carbon governance performance of each province in China? This paper selects 31 provinces in China, constructs a corresponding performance evaluation index system, and further analyzes the low-carbon governance performance of local governments systematically from both static and dynamic aspects using a modified DEA model and Malmquist index. The findings of the study have important policy implications for government environmental governance. Practically, the results indicate that provincial governments cannot rely entirely on increasing the scale of environmental protection investment to improve the efficiency of low-carbon governance but should instead adopt a series of measures and approaches to stimulate public awareness of environmental protection, encourage social forces and enterprises to participate in environmental governance, and form a collaborative development pattern of individual and collective protection. Currently, there is still a need to continue to increase the investment in environmental governance, improve the environmental protection policy system, and improve efficiency and completion of policy implementation. Theoretically, this paper constructs a methodologically revised DEA model and also more accurately estimates the efficiency of low-carbon governance. In addition, this paper is also the first paper to evaluate the performance of various local governments nationwide, and the study fully demonstrates the regional imbalance and inadequacy of low-carbon governance. Of course, there are some shortcomings in this paper. First, although the selection of indicators in this paper is based on a large amount of literature, some important factors may be overlooked, and a more comprehensive evaluation of the government's low-carbon governance performance cannot be made. Second, the time span of the paper is only 6 years, and environmental governance is a long-term task, and analyzing low-carbon governance performance from a longer time span may lead to more findings. Finally, the DEA model used in this paper still has room for improvement, such as a DEA model that integrates environmental efficiency and economic efficiency.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Yu, X. Yang, and K. Li, "Effects of the terms and characteristics of cadres on environmental pollution: evidence from 230 cities in China," *Journal of Environmental Management*, vol. 232, pp. 179–187, 2019.

[2] X. Wang and P. Lei, "Does strict environmental regulation lead to incentive contradiction? - evidence from China," *Journal of Environmental Management*, vol. 269, Article ID 110632, 2020.

[3] S. Yue, R. Lu, H. Chen, and J. Yuan, "Does financial development promote the win-win balance between environmental protection and economic growth?" *Environmental Science and Pollution Research*, vol. 25, no. 36, Article ID 36438, 2018.

[4] X. Liang, M. Ye, L. Yang, W. Fu, and Z. Li, "Evaluation and policy research on the sustainable development of China's rare earth resources," *Sustainability*, vol. 10, no. 10, p. 3792, 2018.

[5] M. Wang and C. Feng, "The win-win ability of environmental protection and economic development during China's transition," *Technological Forecasting and Social Change*, vol. 166, Article ID 120617, 2021.

[6] C. Lu, D. Wang, H. Li, W. Cheng, X. Tang, and W. Liu, "Measurement of the degree of coordination in regard to carbon emissions, economic development, and environmental protection in China," *Applied Sciences*, vol. 11, no. 4, p. 1750, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9768928, 1 page
https://doi.org/10.1155/2023/9768928

Retraction

Retracted: The Application of PSO-SVM Algorithm in the Evaluation System of Sports Competition Events

Security and Communication Networks

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Zhang, "The Application of PSO-SVM Algorithm in the Evaluation System of Sports Competition Events," *Security and Communication Networks*, vol. 2022, Article ID 6865425, 8 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6865425, 8 pages
https://doi.org/10.1155/2022/6865425

Research Article

The Application of PSO-SVM Algorithm in the Evaluation System of Sports Competition Events

Zhenglei Zhang

School of Physical Education and Health, Heze University, Heze, Shandong 274000, China

Correspondence should be addressed to Zhenglei Zhang; zhangzhenglei@hezeu.edu.cn

Received 17 February 2022; Revised 4 March 2022; Accepted 19 March 2022; Published 28 April 2022

Academic Editor: Chin-Ling Chen

With the development of society and the improvement of residents' living standards, sports competition events have attracted much attention as a large-scale event that is popular with the people. Based on the process of holding sports competitions, this paper obtains 14 evaluation indicators from four dimensions of policy, social environment, economy, and natural environment, uses the PSO-SVM algorithm for a comprehensive evaluation, and finally selects Henan Province as the empirical research object. Empirical research shows that the PSO-SVM algorithm has certain applicability in evaluating SCE and can provide a reference for the government and event organizers. The evaluation index system and evaluation model constructed in this paper have certain practical value, but the training data are insufficient and the model still needs to be further improved.

1. Introduction

With the steady and rapid growth of the national economy and the gradual improvement of people's consumption level, sports competition events (SCEs) are becoming more and more frequent and their influence has become wider [1]. Paying attention to sports competitions has become one of the important contents of public cultural life. As the economic benefits and social value of sports competitions are more and more recognized by the public in the modern society, the service pattern of sports competitions needs to be further opened. SCE is not only a project to cultivate competitive talents but also an important means to realize national sports [2]. SCE is a systematic project with large investment, high output, and long-lasting operation. At the same time, the event can attract a large number of domestic and foreign audiences during its holding period, which is also a huge boost to the catering industry, tourism industry, retail industry, transportation industry, and communication industry [3]. The evaluation of SCE needs to be regarded as a project, and its measurement standard is not limited to economic benefits but also comprehensively the social benefits and resource consumption of the project, as well as the subsequent environmental protection. The evaluation of SCE needs to be based on a comprehensive index, which should be used as a reference for future SCE.

In terms of the evaluation indicators of SCE, scholars mostly evaluate sports competitions from the level of the Olympic Games. Xu et al. [4] believe that the sustainable development strategic framework can create infrastructure, ecological environment, organizational governance, and culture for Olympic cities. Gordziejko [5] pointed out that the infrastructure built during the London Olympics brought new vitality to London's tourism industry and provided a reference for its low-carbon lifestyle, which in turn promoted the development of the city. Shin [6] pointed out that large-scale sports events such as the Olympic Games demonstrate the ability of cities to promote urban development through positive economic impacts, including the image of the city or country, the increase in employment opportunities, and the rapid development of tourism. At the same time, it also promotes the renewal progress of the host city, improves the living environment and aesthetic appearance, and achieves sustainable economic growth [7]. The construction demand of the venue is one of the important contents of the overall planning of the Games [8]. Therefore, in the SCE evaluation index, the policies, economy, and ecological environment related to the venue construction are very important [9]. In the early stage of SCE, it is necessary to coordinate the venue and urban development planning and use the venue after the competition as a catalyst for urban transformation. Urban construction can also achieve improvements in infrastructure, housing, and cultural space during the construction of the venue [10]. It can be seen that the evaluation of SCE needs to be carried out from the early, middle, and late stages of the activity, and the evaluation direction should include the future development policy of the city, economic development, and ecological environment protection.

The traditional evaluation methods all use linear regression or estimation methods of descriptive statistics that can be observed at the sample level and use traditional statistical techniques to conduct investigations, including univariate statistical methods, factor analysis, discriminant analysis, and Logit model to achieve evaluation normality and independence of indicators [11]. Scholars have also carried out some research work on sports-related resource optimization and sports project evaluation. Li et al. [12] used the fuzzy evaluation method to evaluate the situation of physical education and used social network and other resources to achieve effective sharing. Zhang et al. [13] used random forest and decision tree algorithms to evaluate related attributes such as school sports information and student learning data. Most of the above evaluation methods evaluate a certain research object from a subjective perspective, and the accuracy and scientificity need to be improved. This paper analyzes the principle and characteristics of the SVM algorithm through the retrieval of the relevant literature [14, 15] and decides to use the PSO algorithm to optimize the SVM algorithm [16].

This paper takes the policy, society, economy, and ecological environment of the SCE host city as the first-level evaluation index and constructs the second-level evaluation index, respectively, fully considering the current situation and sports development of the SCE host city. According to the social development situation, the SCE evaluation index system is constructed, the evaluation results are divided into four grades, and the PSO-SVM algorithm is used for empirical research. The overall evaluation results provide suggestions and references for the holding of SCE.

2. Evaluation Index System of SCE

Project evaluation is a systematic project, and its evaluation index system includes social equity, social justice, mutual adaptability, and sustainable development between the project and the region. When determining the host city of the Olympic Games, the International Olympic Committee needs to accurately and comprehensively analyze the overall impact of the Olympic Games on the host city and even the host country in four aspects: environment, society, culture, and economy. When evaluating sports competitions, it is necessary to start from the early, middle, and late stages of the activity. Sports competitions can not only bring wonderful performances to the audience but also bring considerable economic benefits and the development of related industries. The evaluation of sports events should comprehensively consider economic benefits, resource conservation, and environmental protection. Therefore, the evaluation principles of sports events should include systematic, objectivity, efficiency, and standardization. The development and operation of large-scale sports competitions must follow the viewpoint of system theory, so the evaluation should be carried out from four aspects: policy environment A_1, social environment A_2, economic benefit A_3, and ecological environment A_4.

The index system of SCE is to concretize and use various indexes, so that the evaluation can be measurable. Sports evaluation is not a single-factor judgment work; it is multidimensional, and it is a process in which multiple factors interact and influence each other. After the initial determination of the evaluation indicators, the Delphi method was used many times to modify and improve and it was clarified that the first-level indicators of multimedia classroom teaching evaluation included policy environment evaluation, social environment evaluation, economic benefit evaluation, and ecological environment evaluation. The indicator system is shown in Figure 1, and the detailed explanation is shown in Table 1.

The policy factors of sports competitions are the basic factors of social evaluation. The infrastructure built for the competition project must first meet the development needs of the country and the city and coordinate with economic development. Therefore, the evaluation indicators in terms of policies include the degree of conformity with national guidelines and policies A_{11}, the degree of conformity with urban construction needs A_{12}, and the degree of conformity with regional economic development A_{13}.

The social environment evaluation of sports competitions is an important part of the comprehensive evaluation. According to the project evaluation theory, the evaluation of sports competitions should not only consider the macronecessity of project development but also consider the necessity of onlookers. As an important project of a country or a city, large-scale sports events must be considered from a macroperspective and microperspective and the goals of economic growth and fair distribution must be achieved through the construction of the project. The efficiency goal of sports competition projects requires that investment in various venues and facilities can be increased and national income can be increased. The fairness requirements of sports competitions should necessitate that the increased national income of the project can be reasonably distributed among different income classes, different regions, and investment in consumption. By holding large-scale events, a large number of jobs will be created for different types of labor and more consumption opportunities will be created for the high-income class. Therefore, the evaluation indicators of social links include the employment rate of residents A_{21}, the cultural and entertainment level of residents A_{22}, the population density of the region A_{23}, and the living standards and habits of residents A_{24}.

Economic benefits are the direct impact of sports competitions on social development. The development of sports competitions can effectively promote the development of regional economy, sectoral economy, and national

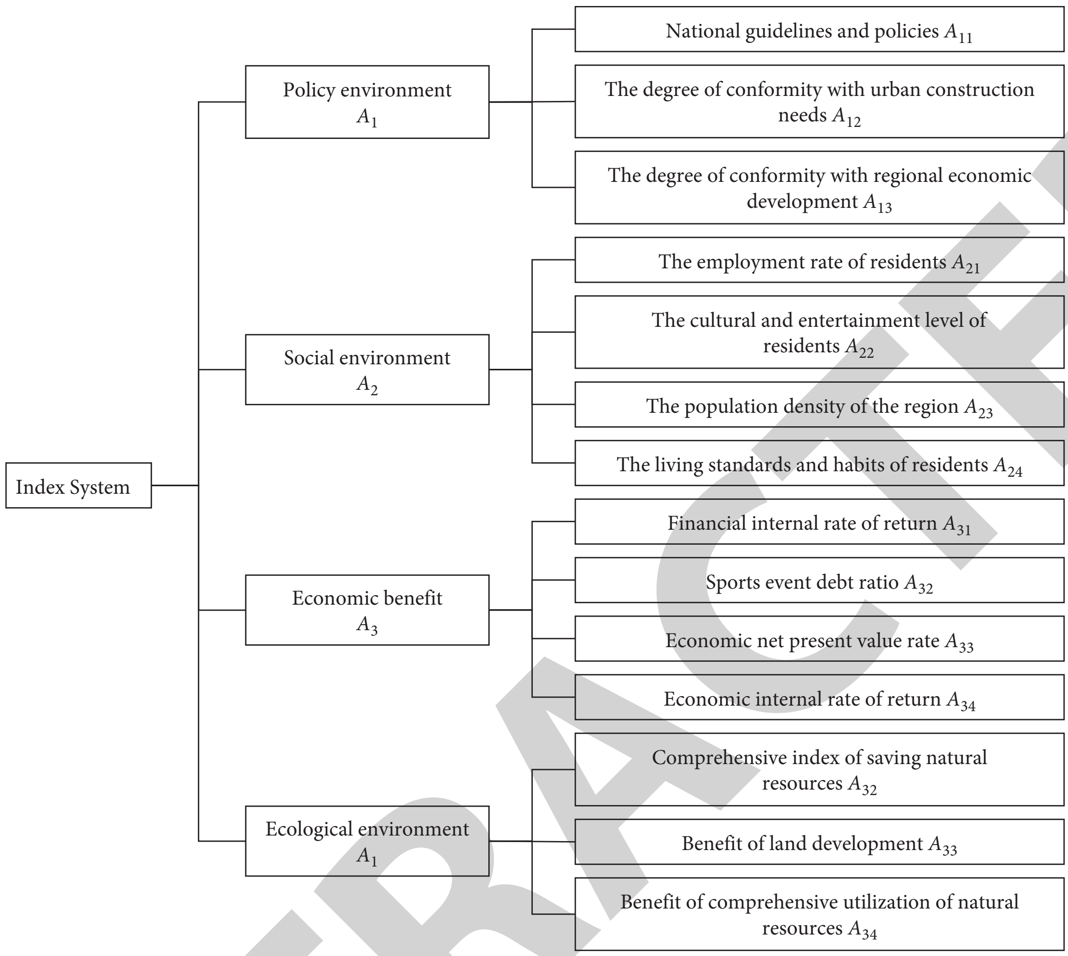


Figure 1: Evaluation index system of SCE.

economy. In terms of economic benefit evaluation, it can be divided into financial evaluation and national economic evaluation. In terms of financial evaluation, it is mainly inspected from financial internal rate of return A_{31} and sports event debt ratio A_{32}. In terms of national economic evaluation, it is based on economic net present value rate A_{33} and economic internal rate of return A_{34}.

Eco-environmental protection is a strategic policy that urban development needs to adhere to all the time. Evaluation of the ecological environment can comprehensively enhance the sustainable development capacity of local areas, provide a clean and neat environment for the development of the project, and ensure that the event will not have a negative impact on the environment during the event. The ecological environment is mainly investigated from three aspects: the comprehensive index of saving natural resources A_{41}, the benefit of land development A_{42}, and the benefit of comprehensive utilization of natural resources A_{43}. The comprehensive index of saving natural resources refers to the need to use the minimum energy to ensure the smooth development of the entire project during the project development process, such as wastewater discharge, sewage treatment, and power consumption. The land development benefit refers to the impact of the project on the surrounding environment and area, that is, whether the project has driven the rational development and utilization of the surrounding land. The benefits of comprehensive utilization of natural resources include the comprehensive utilization of water, electricity, oil, gas, and solar energy.

The above-given evaluation index system can more comprehensively summarize the actual situation in the development of SCE. Taking the evaluation index data of a city's SCE as the input sample of the evaluation model, the evaluation effect of the model can be achieved, and it can provide a reference for the city to host large-scale SCE.

3. The Evaluation Model of SCE

Due to the multifactor coupling and nonlinear characteristics of SCE and its main controlling factors, it is difficult to give accurate results using conventional prediction methods. Therefore, this paper constructs the evaluation model of SCE with the help of the improved support vector machine model with strong robustness and nonlinear mapping ability.

3.1. Evaluation Principle of SVM. Support vector machine (SVM) is a supervised machine learning algorithm. In

Table 1: Explanation of the SCE evaluation index.

Index	Explanation
A_{11} National guidelines and policies	The development of sports economic projects needs to comply with national or regional policies and guidelines, in order to be able to adapt to the development of society
A_{12} The degree of conformity with urban construction needs	Infrastructure construction not only needs to meet the requirements of the competition project but also needs to meet the living needs of the residents after the competition
A_{13} The degree of conformity with regional economic development	Sports competitions need to meet social and economic development and need to meet the needs of residents for the level of economic development
A_{21} The employment rate of residents	Sports competitions will bring more jobs to residents and increase the employment rate of residents
A_{22} The cultural and entertainment level of residents	Residents' consumption level in sports and the popularity of sportsmanship
A_{23} The population density of the region	Population density of cities hosting sports events
A_{24} The living standards and habits of residents	Mainly refers to the residents' physical quality and physical exercise habits
A_{31} Financial internal rate of return	Refers to the discount rate when the sum of the present value of the financial net cash flows of each year of the project is equal to zero throughout the calculation period
A_{32} Sports event debt ratio	The discount rate that makes the economic net present value of a major sporting event equal to zero over the calculation period
A_{33} Economic net present value rate	It is a ratio indicator of the sum of the net present value of the national economy of the investment project and the present value of all investments discounted at the social discount rate
A_{34} Economic internal rate of return	The discount rate at which the cumulative economic net present value over the project calculation period equals zero
A_{41} Comprehensive index of saving natural resources	Refers to the realization of resource conservation and environmental protection by formulating scientific plans during the event
A_{42} Benefit of land development	Benefits of the project to the rational development and utilization of surrounding land
A_{43} Benefit of comprehensive utilization of natural resources	Whether the utilization of resources maximizes the saving of nonrenewable resources

essence, SVM avoids the traditional process from induction to deduction, realizes efficient "transduction reasoning" from training samples to forecast samples, and greatly simplifies the usual problems of classification and regression. The final decision function of SVM is determined by only a small number of support vectors, and the complexity of the computation depends on the number of support vectors rather than the dimensionality of the sample space, which in a sense avoids the "dimensionality disaster." A few support vectors determine the final result, which can not only help us catch the key samples and eliminate a large number of redundant samples but also represent better simplicity and robustness.

SVM transforms the original linear inseparable low-dimensional space training set samples into high-dimensional space G through nonlinear mapping ψ and at the same time constructs the optimal hyperplane to maximize the edge of the classification interval, thereby realizing sample classification. Assuming that within the accuracy ε, the training sample set can be linearly fitted with the following equation:

$$\mathbf{f}(\mathbf{x}) = [\boldsymbol{\omega}\psi(x)] + b, \tag{1}$$

where $\mathbf{x}$ and $\mathbf{f}(\mathbf{x})$ are the input and output vectors, respectively; $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}^T$ and $\mathbf{f}(\mathbf{x}) = \{f_1, f_2, \ldots, f_n\}^T$, in which $x_i \in \mathbf{R}^l$, $f_i \in \mathbf{R}$; n is the number of samples; l is the dimension of the input vector; ω is the weight vector; $\psi(x)$ is a nonlinear mapping; and b is the bias term. In general, the constraints are shown in the following equation:

$$\begin{cases} f_i - \omega\psi(x_i) - b \leq \varepsilon + \mu_i, \\ \alpha\psi(x_i) + b - f_i \leq \varepsilon + \mu_i^*, \end{cases} \tag{2}$$

where μ_i and μ_i^* are slack variables used to coordinate the fitting error and $i = 1, 2, \ldots, n$.

According to the maximum interval principle of SVM, the nonlinear regression problem to be solved is the problem of minimizing the objective function, that is, $\min 1/2\omega^T \cdot \omega + c/n \sum_{i=1}^{n} (\mu_i + \mu_i^*)$, where $c > 0$ is the penalty factor. And, the larger the c, the greater the penalty for misclassified samples. Using dyadic theory, this can be transformed into the corresponding dyadic problem, as shown in the following equation:

$$\max\left[-\frac{1}{2}\sum_{i,j=1}^{n}(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*)(\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^*)K(x_i, x_j) - \omega\sum_{i=1}^{n} f_i(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*)\right], \tag{3}$$

where α_i and α_i^* are the l-dimensional support vectors to be solved, and the constraint is $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$, α_i, $\alpha_i^* \in [0, c]$. In addition, $K(x_i, x_j)$ is the kernel function.

The core of the support vector machine prediction model lies in the selection of the kernel function, and the more commonly used radial basis kernel function (RBF) is chosen in this paper. At this point, the pairwise problem becomes a quadratic programming problem and the optimal solutions α_i and α_i^* can be obtained by solving the quadratic programming problem. Therefore, the linear regression function model can be expressed as the following equation:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b. \quad (4)$$

When using SVM of radial basis kernel function to evaluate SCE, it is necessary to give the optimal penalty factor c and kernel function parameter g, which is the focus and difficulty of using the SVM model for prediction.

3.2. Evaluation Principle of PSO Algorithm. The particle swarm optimization (PSO) algorithm is proposed based on the characteristics related to predation of group animals such as birds and fish and has a strong parameter-finding capability. The PSO algorithm uses massless particles to model social survival rules. Each particle represents an individual, and each particle shares with the whole particle swarm after searching the optimal position in the neighborhood, the optimal particle is used as the local optimal solution of the whole particle swarm, and the optimal solution of the whole particle swarm is obtained after iterating T times. The algorithm is simple and fast to solve and is very suitable for the optimization of complex solving processes such as SCE evaluation. In order to overcome some limitations of the SVM model, the PSO algorithm is chosen in this paper to accelerate the merit-seeking parameters, make the evaluation process more efficient, and enhance the objectivity of the evaluation results.

Suppose that in the D-dimensional search space, m particles form the population $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_k]$. Each particle represents the position in the D-dimensional search space position in the space, which is also a potential solution. Among them, for the u-th particle $\gamma_u = [\gamma_{u1}, \gamma_{u2}, \ldots, \gamma_{u\,D}]^T$, the corresponding fitness value is calculated according to the fitness function and its speed $V_u = [[V_{u1}, V_{u2}, \ldots, V_{u\,D}]^T$. Individual extreme value is $P_u = [P_{u1}, P_{u2}, \ldots, P_{u\,D}]^T$. The global extreme value is $P_g = [P_{g1}, P_{g2}, \ldots, P_{g\,D}]^T$. In the process of iteration, the particle updates its velocity and position by individual and global extremes, as shown in the following equations:

$$V_{ud}^{pp+1} = wV_{ud}^{pp} + c_1 r_1 \left(P_{ud}^{pp} - \gamma_{ud}^{pp}\right) + c_2 r_2 \left(P_{gd}^{pp} - \gamma_{ud}^{pp}\right), \quad (5)$$

$$\gamma_{ud}^{pp+1} = \gamma_{ud}^{pp} + V_{ud}^{pp+1}, \quad (6)$$

where pp is the current iteration number; $u = 1, 2, \ldots, k$; c_1 and c_2 are acceleration factors, which are nonnegative constants; r_1 and r_2 are random numbers distributed between [0, 1]; and w is the inertia weight.

The basic flow of the PSO algorithm is as follows:

Step 1: initialize the population, set basic parameters such as maximum genetic times T, inertia weight w, and acceleration factors c_1, c_2, etc.

Step 2: calculate the fitness function of the particle based on the objective function, and calculate the local optimal solution of the particle population for the current case.

Step 3: optimize the velocity and position of the particles in the population according to the local optimal solution.

Step 4: determine whether the termination condition is met, and terminate the iteration if the requirement is met; otherwise, return to Step 2.

3.3. The PSO-SVM Evaluation Model for SCE. The values of the two parameters c and g in SVM are very important. Here, c is the penalty coefficient; too high a value of c tends to overfit; too small a value of c tends to underfit. g is a parameter that comes with RBF as the kernel function. The size of the kernel parameter g is inversely proportional to the number of vectors supported by RBF. On this basis, this paper takes the influencing factors of SCE as the input vector of the evaluation model; with the ability of PSO algorithm to constantly update and self-learn, the optimal penalty factor and kernel function parameters of SVM are obtained and the SCE evaluation model is constructed, so as to effectively select SCE. In order to predict the classification more accurately and shorten the training time of the model, the K-CV method is used for cross-validation to obtain the optimal c and g. The accuracy in the sense of the training set is used as the fitness function in the PSO algorithm, as shown in Figure 2.

4. Experimental Results and Analysis

In this paper, the SCE of large games in Henan Province is selected as the research object, the basic data collection is carried out according to the SCE evaluation index system summarized above, the actual development of SCE of large games and the evaluation of SCE by relevant personnel are analyzed, and the selectable degree of SCE and 200 data samples can be obtained for model construction and testing. For the convenience of computing, we divided the overall data into 8 datasets, each containing 25 experimental data.

To avoid blind search and time-consuming efforts, this paper places certain restrictions on the relevant parameters of the PSO-SVM model for SCE evaluation. The penalty factor c is taken in the range of [0,100], and the kernel function parameter g is taken in the range of [0,1]. The initial population size is 20, the initial global search ability parameter is 1.5, the initial local search ability parameter is 1.8, and the number of iterative evolutions is 2000. Through model training, the optimal penalty factor is 1.242 and the optimal kernel function parameter is 0.337.

To verify the reasonableness of the PSO-SVM model proposed in this paper for SCE evaluation, we choose SVM without the optimization algorithm and the commonly used machine learning method random forest (RF) model for accuracy comparison. All data will be used for each evaluation method, and the final accuracy comparison results are shown in Table 2 and Figure 3.

The results in Table 2 and Figure 3 demonstrate the superiority of the data in this paper, with average accuracies of 95.84%, 90.49%, and 89.05% for the three methods, respectively. The accuracy rates of the three methods are all above 85%, which is quite convincing. But in comparison, SVM is less stable. In addition, the optimization effect of the PSO algorithm for SVM is well reflected, and the average

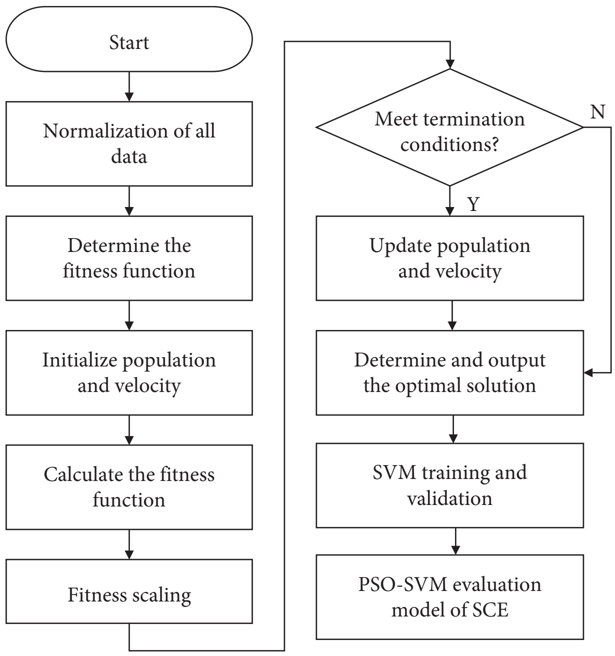


Figure 2: Calculation process of the PSO-SVM evaluation model.

accuracy rate is increased by 5.35%. The evaluation accuracies of SVM and RF are similar, and PSO-SVM has a greater advantage over them.

Figure 4 reveals the difference in the total training plus testing time for the different evaluation methods. The overall difference in runtime between RF and SVM is not significant, with RF running slightly faster by 3.6 ms on average over the 8 datasets. The use of PSO makes the SVM run 12.2 ms faster on average, which means the efficiency is increased by almost 55.2%. This is a good indication of the excellent optimization capability of PSO and also shows that in the sports industry, much research can be done on the basis of SVM to obtain evaluation models with higher accuracy.

The SCE evaluation model in this paper can provide some reference in event selection for the 14th Provincial Games to be held in Henan in 2022. Through research and interviews with people related to the event holding, this paper evaluates some SCEs that are widely popular and vastly discussed. We use the equal-width method to determine the final score of sports events, with a minimum score of 1 and a maximum score of 5. Finally, we divide the ratings into a total of four grades (C, B, A, and S), and S on behalf of this event is very much recommended to choose. The scores of the ten major sports related to the provincial games calculated by the PSO-SVM method are shown in Table 3 and Figure 5.

Overall, due to economic development and the enthusiasm of the people, the competitive sports environment has flourished and many sports have broken through the limitations of venues and gained the support of more citizens. Badminton and table tennis, with their wide participation and high popularity, are rated S and are very suitable for SCE. Track and field, wrestling, and karate are not as popular

Table 2: The accuracy of different evaluation methods.

Dataset	Accuracy (%)		
	PSO-SVM	SVM	RF
1	97.49	93.21	90.34
2	96.82	92.33	88.53
3	96.45	93.19	89.23
4	93.97	91.05	87.44
5	94.67	89.33	90.25
6	95.66	87.82	89.32
7	95.48	89.57	87.32
8	96.15	87.42	89.97

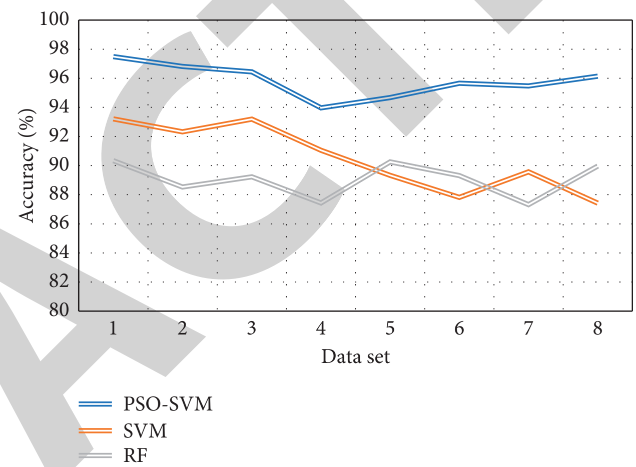


Figure 3: The accuracy of different evaluation methods.

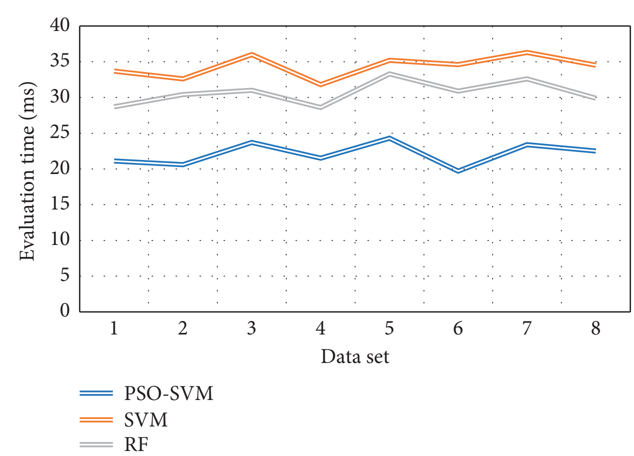


Figure 4: Comparison results of training time of different models.

Table 3: SCE rating scale.

	Track and field	Badminton	Table tennis	Tennis	Swimming
Grade	A	S	S	B	B
	Rowing	Wrestling	Triathlon	Karate	Skateboard
Grade	A	A	B	A	A

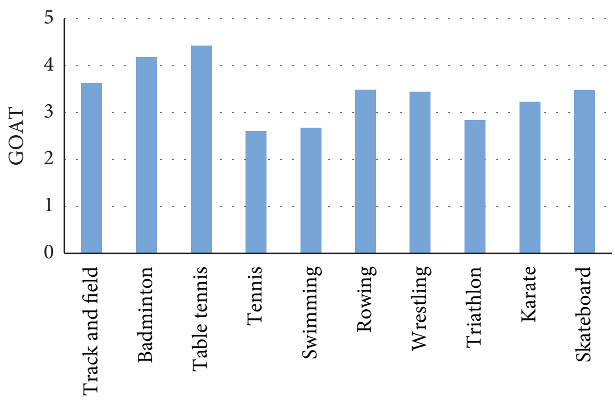


FIGURE 5: SCE rating scale.

as the above ball sports but require fewer demanding venues and have a higher number of participants, thus receiving an A rating. Tennis, swimming, and triathlon, on the other hand, require a high level of human physical fitness and have more restrictions on venues and equipment. According to the evaluation index system proposed in this paper, they are not well suited to the lifestyle habits of local residents, resulting in a B rating. However, these sports are highly spectator-friendly and have a certain economic potential. Their development in the local community can be supported by improving publicity and creating a better economic environment.

5. Conclusion

In order for the public to actually feel the charm of competitive sports, it is necessary to scientifically select the projects of the major games. There are many factors affecting SCE selection, complex sources and complex nonlinear relationships between factors and evaluation results, so this paper investigates the PSO-SVM evaluation model to accurately characterize SCE. In this paper, a comparative experimental approach is used, and the results show that RF and SVM have similar evaluation results, while PSO improves the accuracy of SVM by 5.35% and speeds up the running time by 55.2%. Therefore, it also demonstrates that the evaluation model in this paper is sufficiently convincing. The evaluation model is applied to the SCE selection for the Henan Provincial Games and yields a rating of S for badminton and table tennis, which are by far the highest rated sports. However, swimming, tennis, and triathlon all have a lot of room for improvement, which confirms the generalizability of the model.

A good model can help the evaluation to be carried out with half the effort, and now, there are many optimization algorithms and a lot of research on SVM models. We hope to combine more cutting-edge algorithms afterwards to bring the model accuracy to a higher level and be able to put the scale of sports events to the whole country.

Data Availability

The labeled dataset used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the Heze University.

References

[1] W.-K. Kim, "A study on the development of expertise the instructors to promote life sports," *Journal of the Korea Entertainment Industry Association*, vol. 11, no. 6, pp. 153–163, 2017.
[2] Y. Ma and M. Kurscheidt, "The National Games of China as a governance instrument in Chinese elite sport: an institutional and agency analysis," *International Journal of Sport Policy and Politics*, vol. 11, no. 4, pp. 679–699, 2019.
[3] K. Bodin and L. Misener, "Non-mega sport events as sociopolitical projects: a case study of the Canada Games," *International Journal of Sport Policy and Politics*, vol. 12, no. 4, pp. 513–529, 2020.
[4] Z. Q. Xu, B. L. Sun, and X. Y. Dong, "Exploration on the realization of the urban heritage vision of beijing winter Olympics under the "sustainability strategy framework"," *Journal of Chengdu Sport University*, vol. 46, no. 4, pp. 89–94, 2020.
[5] T. Gordziejko, "The London 2012 Olympic legacy and the wonder factor: implications for culture and the intrinsic versus instumental debate," *Journal of Policy Research in Tourism, Leisure and Events*, vol. 6, no. 1, pp. 80–84, 2014.
[6] H.-b. Shin, "Developing SPORT LEGACY through PostUse of mega-sport event facilities: the case of winter sport," *J-Institute*, vol. 5, no. 1, pp. 17–28, 2020.
[7] H.-b. Kim, "'A study on CRIME prevention and policing in KOREA's urban regeneration," *J-Institute*, vol. 3, no. 2, pp. 6–12, 2018.
[8] H. S. Deng, M. Mancin, and L. Zhang, "Between urban renovation and olympic sporting legacy: the case of shougang," *Movement & Sport Sciences*, vol. 107, pp. 53–65, 2022.
[9] Y. Deng, S. W. Poon, and E. H. W. Chan, "Planning mega-event built legacies: a case of expo 2010 beijing," *Habitat International*, vol. 53, pp. 163–177, 2015.
[10] J. Davis, "'Futurescapes of urban regeneration: ten years of design for the unfolding urban legacy of London's Olympic Games, 2008-2018," *Planning Perspectives*, vol. 34, no. 5, pp. 877–901, 2019.
[11] P. Wicker, J. Prinz, and T. Von Hanau, "Estimating the value of national sporting success," *Sport Management Review*, vol. 15, no. 2, pp. 200–210, 2012.
[12] L. LiY, C. X. Wang, and Z. J. Li, "Optimization algorithm of resource allocation based on fuzzy assessment method," in *Proceedings of the International Conference on Mechatronics Engineering and Computing Technology (ICMECT), Shanghai, PEOPLES R CHINA*, Shenyang, China, November 2014.
[13] Z. F. Zhang, Z. J. Zhao, and D. S. Yeom, "Decision tree algorithm-based model and computer simulation for evaluating the effectiveness of physical education in universities," *Complexity*, vol. 2020, Article ID 8868793, 11 pages, 2020.
[14] Z. Liu, H. Cao, X. Chen, Z. He, and Z. Shen, "Multi-fault classification based on wavelet SVM with PSO algorithm to

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9768082, 1 page
https://doi.org/10.1155/2023/9768082

Retraction

Retracted: Cross-Border E-Commerce Logistics Collaboration Model Based on Supply Chain Theory

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Xie, H. Wang, and J. Jiao, "Cross-Border E-Commerce Logistics Collaboration Model Based on Supply Chain Theory," *Security and Communication Networks*, vol. 2022, Article ID 1498765, 13 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1498765, 13 pages
https://doi.org/10.1155/2022/1498765

Research Article

Cross-Border E-Commerce Logistics Collaboration Model Based on Supply Chain Theory

Cuijie Xie, Haijuan Wang, and Jianhong Jiao

Department of Logistics Management, Hebei Jiaotong Vocational and Technical College, Shijiazhuang 050035, Hebei, China

Correspondence should be addressed to Jianhong Jiao; jiaojianhong@hejtxy.edu.cn

Received 22 February 2022; Revised 7 March 2022; Accepted 21 March 2022; Published 28 April 2022

Academic Editor: Chin-Ling Chen

With the rapid development of cross-border e-commerce, however, logistics has become a bottleneck in the development of cross-border electricity traders. The results of studies on cross-border e-commerce logistics are still less, and the relevant theoretical studies are not yet mature enough. As cross-border e-commerce occupies a share of foreign trade in foreign trade increases, so does their influence. In order to eliminate bottlenecks in the cross-border logistics of an electric enterprise, it is of great importance to systematically study the issues of synergy in the logistics of the supply chain of a cross-border electric enterprise and validate how cross-border traders and cross-border logistics work together using cross-border discussion based on the perspective of cross-border e-commerce ecosystem. At the same time, an analysis of the need for cross-border logistics collaboration electricity traders and cross-border logistics is being carried out, as well as an in-depth study of synergy mechanisms between cross-border electricity traders and cross-border logistics based on a cross-border ecosystem perspective. The empirical results show that cross-border logistics is available function service capability; cross-border logistics information sharing level, cross-border logistics resource optimization and allocation capability, and the opening level of cross-border logistics environment have different contributions to the impact on the efficiency of cross-border e-commerce logistics. Among them, the level of cross-border logistics information exchange has the most significant influence on the logistics efficiency of cross-border traders, followed by cross-border logistics functional services capabilities, again the level of openness of the cross-border logistics environment, and finally, the ability to optimally distribute cross-border logistics resources.

1. Introduction

In recent years, China's e-commerce has maintained a rapid growth momentum, and online shopping has fundamentally changed people's concept of consumption and consumption habits and has completely entered people's life. Domestic e-commerce enterprises have also achieved a huge leap in their transaction volume and fully demonstrated the great potential of e-commerce development. Under the situation that the e-commerce industry has become the engine driving the rapid and sustainable growth of the national economy and expanded rapidly, and the in-depth development of electronic information technology and economic globalization, online transactions have also gone abroad. Therefore, cross-border e-commerce came into being and had been associated with the general and with the trends of power companies [1]. Figure 1 shows the flowchart of cross-border trade and logistics. The rapid developing cross-border e-commerce is accompanied by the quiet formation of cross-border merchant logistics and the formation of several characteristics that differ from traditional cross-border logistics. Cross-border Electronic Commerce accelerates the process of electronification and informatization of traditional trade. Electronation is that the Internet has changed the information exchange mode of transaction subjects between different customs territories. For example, through the cross-border Electronic Commerce electronic platform, buyers and sellers can directly know each other's information and interact with each other directly. Informatization means that the transaction behavior between transaction subjects in different customs areas should be supervised within a certain range, such as the seller's delivery punctuality, logistics cycle, logistics security, and buyer's evaluation. Although e-commerce relying on the Internet

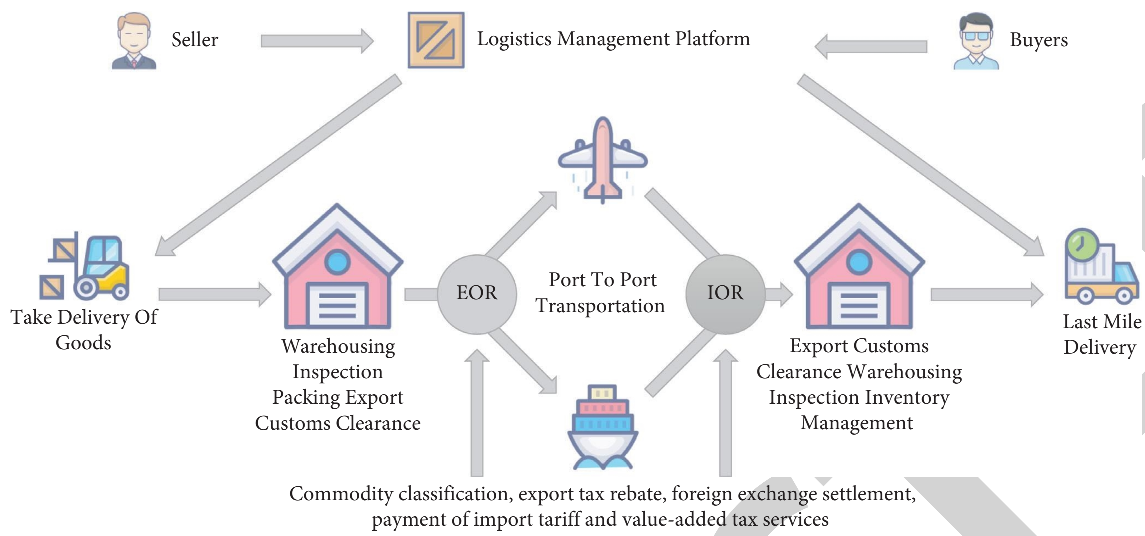


Figure 1: Cross-border e-commerce logistics map process.

has expanded the capacity of information in the short term, a variety of information from both buyers and sellers has been collected on the cross-border e-commerce trading platform, and this information has increased sharply with the increase of public awareness. In this case, the cross-border at the virtual end will develop rapidly, while the cross-border Electronic Commerce supporting logistics at the physical end is difficult to upgrade in a short time. This is because there are multiple contradictions in cross-border e-commerce logistics and logistics at the physical end: domestic logistics and international logistics, logistics between countries, low-end logistics and high-end logistics, mass market, and niche market [2].

2. Literature Review

As early as the 1960s, the term "supply chain" was proposed. Supply chains are the process of transferring the purchased raw materials to users through production, processing, and sales. Matsui K. and others pointed out to a model of e-commerce + supply chain, including cross-border procurement, processing, distribution, return, and other basic contents of goods, which is the future development of cross-border e-commerce [3]. Guan S. and others defined the trade subjects of cross-border Electronic Commerce supply chain and believed that these cross-border trade subjects mainly include overseas purchasers, manufacturers, international e-commerce platform, logistics enterprises, customs, consumers, etc. [4]. Faghat E. and others defined cross-border Electronic Commerce supply chain as cross-border e-commerce using the supply chain to carry out cross-border electronic transactions, cross-border logistics, cross-border supply, and other activities, to connect suppliers, customs, logistics providers, and end consumers into an integral functional network chain [5]. Baier C. and others studied the risk of cross-border e-commerce supply chain using structural equation model and showed that the cross-border e-commerce supply chain is more complex than the traditional supply chain, and the cross-border e-commerce supply chain is a mesh supply chain. The chain from merchants to buyers has weak flexibility, while the emergence of mesh supply chain increases the elasticity of supply chain [6]. Perdana Y. R. and others show that the traditional supply chain is composed of four parts: suppliers, suppliers, distributors, and users. Logistics, capital flow, and information flow will occur. The surface structure of cross-border e-commerce supply chain is simpler, which is composed of manufacturers, e-commerce platforms, and consumers [7]. Zatta F. and others believe that, compared with the traditional supply chain, the current cross-border e-commerce supply chain shortens the supply chain with the help of cross-border e-commerce platform, greatly reduces the cost, and makes the transaction more global, transparent, and timely with the help of the Internet. Cross-border e-commerce supply chain performance can affect the operation results of cross-border e-commerce. Improving supply chain performance can improve the competitiveness of the supply chain and enhance the profit margin of enterprises [8]. Zhang X. and others believe that the e-commerce supply chain is the sum of the values reflected in the operation process and results of enterprises. In recent years, the research on the performance evaluation system of cross-border e-commerce supply chain generally adopts the Balanced Scorecard (BSC) to establish the performance evaluation index system and uses the interpretive structure model (ISM) to decompose the index system at multiple levels, to analyze how to improve the performance level [9]. Lopatin A. and others have shown through research that the factors affecting cross-border e-commerce supply chain performance have seven characteristics, among which the informatization level is the lowest influencing factor, crisis response ability, logistics ability, innovation ability, customer satisfaction, and sustainable development ability are distributed at the middle level, and the top influencing factors are sales and profits [10]. Tarigan Z. and others established the system of cross-border e-commerce supply chain from five aspects: customer, finance, operation process, growth ability, and information technology, and analyzed the influencing factors by using the explanatory structure model. It is found that the influencing factors can be divided into nine levels. Overall, growth capability and information technology are the basic layer, the operation process is the operation layer, and customers and finance are the result layer. In the basic layer, information technology is the core competitiveness of the development of cross-border e-commerce enterprises. The growth ability is the basis for the sustainable development of enterprises. The development of enterprises should start from the basic layer. The operation layer should pay attention to improving the logistics capacity, to improve the supply chain performance and promote the development of cross-border e-commerce [11]. After summarizing the characteristics of BSC and ISM, Guo, J. and others used the direct weight method to determine the weight of the evaluation index system. Then based on the evaluation method, they evaluated the supply chain performance from five aspects: customer satisfaction, finance, internal process, future development, and social responsibility, conducted empirical analysis, and gave suggestions to cross-border e-commerce enterprises [12].

3. Method

3.1. Joint Communication and Mechanism of Cross-Border Electricity Traders and Cross-Border Logistics

3.1.1. Index Selection. This study intends to conduct an empirical study on the synergy between data on the development of a cross-border electric enterprise and data analysis of a cross-border logistics industry. About the choice of indicators, there are currently relatively few examples of direct statistics on cross-border traders and cross-border logistics, to improve the targeting and validity of research results, based on relevant research literature references, as well as relevant staff consultation, ultimately indicators of the level of development of cross-border traders. As a representative indicator to measure the development level of the cross-border electric enterprise, the total volume of cross-border business transactions (KJDS) is chosen, which directly responds to the current state of development of the cross-border electric enterprise in our country in terms of economic level. The cross-border logistics system mainly includes two pillar subsystems of transportation and storage, of which the cost of transportation subsystem accounts for the vast majority. Therefore, three indicators such as the repair system in the garden (TTL), International Corporation/Hong Kong, Tometi Express volume (KDL), and international/Hong Kong, Macao, and Taiwan express business revenue (KDR) are selected to measure the development level of cross-border logistics [13].

3.1.2. Empirical Research. This article analyzes association of traders and cross-border associated logistics calculation coefficients between variables. The study has a high positive correlation between the throughput of foreign trade cargo of the country's ports, the volume of international courier operations, the income from international courier operations, and the total volume of cross-border transactions with electric enterprises, assuming that h1 is created. There is a high positive correlation between the variables, but perhaps this is actually the phenomenon of "pseudo-regression" of changes in the same trend caused by some special economic phenomenon. A common and reliable way to identify "false regression" is to determine it from its roots, i.e., by testing for time series stability. Thus, to make the effectiveness of the subsequent analysis in this article, you must first check the stability of the various relevant data. In this article, the ADF method performs a single root control of four variables LNKJDS, LNTTL, LNKDL, LNKDR and their difference sequences. It is known from the control results that, at a significance level of 5%, four variables LNKJDS, LNTTL, LNKDL, LNKDR belong to an unstable time series, a unit root, the same first difference. Corresponding to it is also a transient time series at a significance level of 5%. But in the two phases, the statistical ADF values are smaller than the critical ADF value at the 5% level, while the second-order differential values of these four variables become smooth time series. Thus, the four variables LNKJDS, LNTTL, LNKDL, and LNKDR belong to a single-row sequence of the second order and satisfy the conditions of subsequent cointegration control and causal testing. Although the four variables LNKJDS, LNTTL, LNKDL, and LNKDR are nonstationary sequences, they all meet the second-order single integration; it can be considered that the linear combination between these four variables may be stationary, and the cointegration relationship test can be further used to judge their long-term stable equilibrium relationship and then test. For cointegration tests, the EG method and the Johansen method are commonly used. With limited quantity variable samples, this paper uses the EG control method to control the cointegration ratio. The EG method is based on a covariance test of the regression residual, with autovariables and cointegration relationships between variables, i.e., self-diagnosis discrimination and smoothing of the residual sequences of the variable regression equation. To test for residual sequence stability, the ADF test method is usually used, and ADF critical values cannot directly take on the critical values obtained by the EViews software, but special tables of critical values compiled by Engle-Granger for auxiliary judgment must be used [14]. Since LNKJDS, LNTTL, LNKDL, and LNKDR are nonstationary sequences of I (2), the following coordinated regression can be performed:

$$\begin{aligned} LNKJDS_t &= \underset{(3.25961)}{1.725493}\, LNTTL_t + \underset{(18.24582)}{1.296548}\, LNKDL_t \\ &\quad + \underset{(-3.269851)}{-1.068592}\, LNKDR_t + \varepsilon_t R^2 \\ &= 0.965834;\, S.E. \\ &= 0.152698;\, DW \\ &= 2.682942. \end{aligned} \tag{1}$$

Each variable in the formula of the regression (1) satisfies the test at the 5% significance level, and the degree of fitting between the calculated and actual values is also very good. Check the cointegration relationships between the four variables above, i.e., the need to discriminate against the stability of the residual sequences in formula (1) of the cointegration equation, that is, to judge whether the sequence $\{\varepsilon_t\}$ is I (0). Since equation (1) does not contain constant term and trend term, it is considered in the residual test regression model. Assuming that the lag order is 1, it can be called AEG test because it is a cointegration test based on residual. AEG regression results are as follows:

$$\begin{aligned}\Delta\hat{\varepsilon}_t &= \underset{(-2.245298)}{-0.126853} + \underset{(2.359248)}{0.035691}\,t - \underset{(-5.638429)}{2.865492}\,\hat{\varepsilon}_{t-1} + \underset{(3.152691)}{0.8125762}\,\Delta\hat{\varepsilon}_{t-1} R^2\\ &= 0.965824;\, DW\\ &= 3.169250;\, S.E.\\ &= 0.0583291.\end{aligned} \tag{2}$$

According to formula (2), AEG = -5.746293, and the regression coefficients of all variables meet the hypothesis of rejecting zero at the significance level of 10%. Because DW = 3.206622, the value of DW is too large, and the residual sequence in formula (2) may contain negative autocorrelation. With the help of the Lagrange multiplier (LM) test, it is known that no significant test results were obtained from the period of delay $P = 1$ to the invalidity of the degree of freedom ($P = 6$), which proves that there is no autocorrelation in the residual sequence in (2) (that is, a lag term added to $\Delta\hat{\varepsilon}_t$ fully meets the dynamic requirements). When $n = 4$, $t = 9$ and there are constant items and trend items, the critical value of MacKinnon bivariate cointegration is $C_{\alpha=0.05} = -6.281857$ at the 5% level, threshold is $C_{\alpha=0.10} = -5.509931$ at the 10% significance level, and $C_{\alpha=0.10} < AEG < C_{\alpha=0.05}$, indicating that $\Delta\hat{\varepsilon}_t$ is I (0), and there is cointegration between the 4 variables, lnkjds, lnttl, lnkdl, and lnkdr, with 10 percent significance level. As can be seen from the above preliminary analysis, it is known that there is stable and balanced relationship between the cross-border electric dealer and the variables of different indicators of cross-border logistics, assuming that H2 is created [15]. The absolute value of the coefficient in the cointegration regression equation for all three indicators of cross-border logistics is greater than 1, which indicates that, in terms of long-term relationships, it has a slightly greater impact than on cross-border logistics. It should be noted that the LNTTL and LNKDL coefficient is positive, while the LNKDR coefficient is negative, that is, the throughput of foreign trade cargo of national ports and the volume of international courier transactions have a positive incentive effect on the total volume of cross-border electricity commercial transactions, and receipts from international courier operations have a certain retroactive deterrent effect on the overall volume of cross-border commercial transactions in electricity, describing that the volume of cross-border logistics operations is changing in the same direction as the total volume of cross-border traders, while the increase in income from cross-border logistics operations (for example, the increase in logistics prices, cross-border operations) hinders the development of cross-border traders to a certain extent. Among them, the largest load in total cross-border e-commerce transactions is exerted by the nation port of foreign trade cargo turnover. The specific efficiency is that the LNTTL regression coefficient in the cointegration regression equation is 1.704896, which is much larger than the other two index coefficients. In general, the long-term relationship between cross-border electricity consumers and cross-border logistics manifests itself mainly in positive stimulation of each other, but there are also certain reverse disincentive effects, and cross-border has a relatively greater cross-border impact on electricity consumers than on cross-border transportation.

3.1.3. Granger Causality Test. Although the results of cointegration test show that there is a long-term stable equilibrium relationship between variables, it does not explain the cause and caused relationship between variables, so it needs to be verified by the Granger causality test. Conduct a Granger causal test on the total amount of cross-border transactions with electric enterprises of variables representing the development of cross-border merchants, as well as on the total throughput of foreign trade expenses of the ports of the country with variables representing cross-border logistics, the volume of transactions of international/Hong Kong, Australian and Taiwanese couriers, and income from operations of international/Hong Kong and Taiwanese couriers, the results of which are shown in Table 1 [16].

The results show that, at 10% significant level, LNTTL is a Granger cause of LNKJDS. What is cross-border logistics? It is a Granger cause for cross-border merchants; LNKJDS is the Granger cause for LNKDL, LNKDR, and this shows that the cross-border electric operator is the Granger cause of cross-border logistics. The reason of the audit was statistically reflected: the growth in the capacity of foreign trade in the country, representing the cross-border logistics, stimulates cross-border development electrical industry; the growth in the total volume of cross-border commercial transactions, representing the development of cross-border electricity traders, also stimulated the growth in the volume of courier transactions and operating income. Cross-border electric merchants logistics are Granger causal relationships that influence each other and develop together in the long term, assuming that H4 will be created.

3.2. Construction of Synergy Evaluation Model between E-Commerce and Logistics. The joint development of traders and logistics refers to the cooperation and mutually reinforcing development between cross-border traders and we will develop cross-border logistics. The levels of synergy between them can be measured by the synergy between cross-border electricity traders and cross-border logistics. Cross-border merchants, as the current trends in e-commerce, develop according to the development level of China's foreign trade. The development of cross-border electrical business has stimulated the progressive cross-border logistics development; at the same time, the development of cross-border energy services and only the joint and orderly development between them could encourage

Table 1: Granger causality test results.

Original hypothesis	Lag order	F statistic	Probability	Conclusion
LNTTL is not Granger cause of LNKJDS	1	13.261	0.0124	Refuse
LNKJDS is not Granger cause of LNTTL		2.5624	0.1539	Accept
LNKDL is not Granger cause of LNKJDS	1	0.00026	0.9826	Accept
LNKJDS is not Granger cause of LNKDL		5.2614	0.0680	Refuse
LNKDR is not Granger cause of LNKJDS	1	0.01521	0.9241	Accept
LNKJDS is not Granger cause of LNKDR		4.26051	0.0937	Refuse

China's cross-border electrical supply chains and cross-border electrical niches to move in a healthier direction. According to current scholars, China's merchants and logistics services are in a low synergy stage, and they will be tested in this article. The question of how to build a measurement model to evaluate the cross-border synergy traders and logistics is also the focus of this article [17]. Thus, in this chapter, a model of cross-border electrical interaction with cross-border logistic will be built in accordance with the following processes:

(1) Establish sequence parameters and subsystems.
(2) Calculate the contribution of order parameters to the subsystem.
(3) Calculate the order degree of subsystem.
(4) Calculate the system synergy.

And on this basis, carry out the calculation of the synergy through China's cross-border electricity traders and cross-border logistics in the past six years and analyze the principle of synergy.

3.2.1. Establishment Principles of Evaluation Index System. We will promote the coordinated development of cross-border e-commerce and cross-border logistics. According to Haken's synergetic theory, the evolution of things is controlled by order parameters, and the final structure and order degree of evolution depend on order parameters. Order parameter is not only the measurement of subsystems, but also the representation of the synergy between subsystems. The essence of increasing synergy between two subsystems is that changes in the order parameters affect the ordering of the subsystem and then the synergy between subsystems. Thus, the size of the "order" parameter can be used to indicate the degree of macrostability of the subsystem, when "order" parameter is zero, the macro-subsystem is disordered, and when the "order" parameter reaches the optimal critical point, a macro-ordered organization process appears in the system. Based on this, this paper assumes that the cross-border e-commerce system is a subsystem. And the cross-border logistics system is a subsystem S_2. The integrated system composed of the two is s, that is, $S = \{S_1, S_2\}$. The coordinated development of S_1 and S_2 or the improvement of the degree of synergy between them is the improvement of the degree of order of system s. Let $e_1 = \{e_{11}, e_{12}, e_{13}, \ldots, e_{1k}\}$ be the cross-border e-commerce order setting and $e_2 = \{e_{21}, e_{22}, e_{23}, \ldots, e_{2k}\}$ be the order parameter of cross-border logistics, where $k = 1, 2, 3, \ldots, n$ represents the number of order parameter indicators in the subsystem. e_{1k} and e_{2k} are the order parameter of cross-border e-commerce system and cross-border logistics system, respectively, and their value range is $a_{1k} \le e_{1k} \le b_{1k}$ and $a_{2k} \le e_{2k} \le b_{2k}$, respectively ($b_{mk}$ and a_{mk} are the ideal optimal value and ideal minimum value of e_{mk}). Because this paper studies the degree of synergy between the two subsystems, it is necessary to ensure that the order parameters between the two subsystems can have sufficient correlation. Therefore, the correlation degree of the order parameters will be screened. Chinese mammoth power companies take the road of logistics benchmark started late and the availability of data is low. In order to ensure the accuracy of the experiment, this paper uses the grey correlation analysis method with low requirements on the age of data to screen the indicators. MATLAB14.0 software is used to screen the correlation degree of order parameters. If the correlation degree is greater than or equal to 0.6, it indicates that there is strong correlation between subsystems. When the relevance is strong, sort according to the correlation coefficient and select the order parameter at the top.

3.2.2. Calculation of Contribution Degree of Order Parameter to Subsystem. According to the synergetics theory, the contribution of order parameters to its subsystem can be calculated by the efficacy function. Since this paper only considers the slow order parameters, which has a positive effect on the stability of the system, the calculation formula is as follows:

$$u_m(e_{mk}) = \frac{e_{mk} - a_{mk}}{b_{mk} - a_{mk}} \quad m = 1, 2;\ k = 1, 2, 3, \ldots, n. \tag{3}$$

$u_m(e_{mk}) \in [0, 1]$, The higher the value of $u_m(e_{mk})$ is, the greater the upward contribution of order parameters to the order of the subsystem is, and vice versa.

3.2.3. Calculation of Contribution of Order Parameters to the Whole Cross-Border E-Commerce and Cross-Border Logistics Coordination System. The contribution of the order parameter to the total system is shown by subsystem flowchart integrating the efficacy function value of the order parameter index. The contribution of order parameters to the whole system can be obtained by calculating the weighted average of order parameters.

$$u_m(e_m) = \sum_{k=1}^{n} w_{mk} u_m(e_{mk}), \tag{4}$$

where $w_m = \sum_{k=1}^{n} w_{mk} = 1$, w_{mk} is the weight of the corresponding $u_m(e_{mk})$. The size of w_{mk} can be weighted by critical weighting method. Critical method considers the influence of index transformation size on the weight. Because the samples studied in this paper are time series samples, the deviation weighting method is more objective and comprehensive than the direct weight method. The calculation formula is

$$\begin{aligned} c_{mk} &= \sigma_k \sum_{k=1}^{n} (1 - \rho_{mk}), \\ w_{mk} &= \frac{c_{mk}}{\sum_{k=1}^{n} c_{mk}}. \end{aligned} \tag{5}$$

c_{mk} indicates the influence degree of the k-th index on the whole evaluation index system, σ_k indicates standard deviation of the k-th evaluating indicator, and ρ_{mk} indicates the correlation coefficient between index m and index K, that is, the correlation coefficient obtained from the grey correlation matrix.

3.2.4. Build the Calculation Equation of Cross-Border E-Commerce and Cross-Border Logistics Complementing Each Other. The relationship between traders and logistics is in a process of dynamic change that changes as time changes, making it necessary to dynamically measure the synergy between cross-border electricity traders and cross-border logistics. Assume that, in a specific stage t_0, the order degree of the cross-border e-commerce system and the cross-border logistic system in this period is $u_m^{t_0}(e_m)$. When the system continues to evolve and develop to time t_1, the order degree of the two subsystems S_m is $u_m^{t_1}(e_m)$. If $u_m^{t_1}(e_m) \geq u_m^{t_0}(e_m)$, the whole composite system is in positive synergy in the period $[t_0, t_1]$, indicating that the two have positive synergy, and the degree of synergy is expressed by U; the calculation formula is

$$U = \sqrt[n]{\prod_{m=1}^{n} \left| u_m^{t_1}(e_m) - u_m^{t_0}(e_m) \right|}. \tag{6}$$

The value range of u is [0, 1]. When u is 0, the system is extremely uncooperative, and when u is 1, the system is extremely cooperative.

3.2.5. Construction of Evaluation Model. Intuitively speaking, the coordination of composite system refers to the harmonious coexistence between its constituent subsystems under the action of suborganizations within the system and regulation and management activities from the outside, so providing the information required to obtain all the information effect of the system. The subsystem always has spontaneous and irregular independent motion, and at the same time, it is affected by the joint action of other subsystems; there is a cooperative motion formed by the correlation between subsystems. There are many control parameters in motion, which are divided into "fast" and "slow" variables, and the "slow" parameter is dominant. As the control parameters change, when the system approaches the critical point, the associativity that forms between subsystems gradually increases. When the control parameters reach the "threshold," the correlation between subsystems plays a leading role. Therefore, there is synergy between subsystems determined by correlation in the system, and there is a macrostructure or type [18].

System order degree.

Considers the supply chain system $S = \{S_1, S_2, \ldots, S_k\}$, where S_j is the j-th subsystem compounded into s, and $j = 1, 2, \ldots, k$. The interaction between S_j produces the overall synergy effect of supply chain. Therefore, the supply chain system can be abstractly expressed as $S = f\{S_1, S_2, \ldots, S_k\}$, where f is the composite factor. For the subsystem S_j, $j = 1, 2, \ldots, k$, set sequential parameters $e_j = (e_{j1}, e_{j2}, \ldots, e_{jn})$, e_{ji} is within a certain control range, and the upper and lower limits are β_{ji}, α_{ji}, and $i \in [1, n]$, respectively. Since the change in the sequence parameters affects the system over time, the degree of order of the system increases or decreases, and the degree of order of the system increases as the components of the sequence parameters increase or decrease. Therefore, in the subsystem S_j, do not lose custom; it is assumed that as the values of variables $e_{j1}, e_{j2}, \ldots, e_{jm}$ increase, the degree of system order first increases and then decreases. Thus, the flowchart of the sequential parameter components is defined:

$$u_j^0(e_{ji}) = \begin{cases} \dfrac{e_{ji} - \beta_{ji}}{\alpha_{ji} - \beta_{ji}}, & i \in [1, m], \\ \\ \dfrac{\alpha_{ji} - e_{ji}}{\alpha_{ji} - \beta_{ji}}, & i \in [m+1, n]. \end{cases} \tag{7}$$

In the formula, all variables are taken at t^0 time.

According to formula (8), the larger the data value of $u_j^0(e_{ji}) \in [0, 1]$, the greater the effect of e_{ji} on the order of the system. In addition, it should be noted that, in fact, the value of e_{ji} can be large or small, but if its value is too large or too small it is not appropriate, so its order degree can meet formula (8) by adjusting its upper and lower limits.

The overall effect of e_j on s_j order can be expressed by the integration of $u_j^0(e_{ji})$. The integration calculation of $u_j^0(e_{ji})$ is generally completed geometric mean method or linear weighting method. The system order level of the subsystem at t^0 is

$$u_j^0(e_j) = \sqrt[n]{\prod_{i=1}^{n} u_j^0(e_{ji})}. \tag{8}$$

Or

$$u_j^0(e_j) = \sum_{i=1}^{n} w_j u_j^0(e_{ji}), \quad w_j \geq 0, \; \sum_{i=1}^{n} w_j = 1. \tag{9}$$

In the formula, all variables are taken at t^0 time.

System synergy.

According to the calculation of system order degree, it can be concluded that, at the initial time t^0, the order degree of each subsystem is $u_j^0(e_j)$, while at the current time t^1, the order degree is $u_j^1(e_j)$. Therefore, DSCC is defined as the system coordination degree; then:

$$dscc = \theta\sqrt{\left|\sum_{j=1}^{k}\left[u_j^1(e_j) - u_j^0(e_j)\right]\right|}, \tag{10}$$

where $\theta = (\min_j[u_j^1(e_j) - u_j^0(e_j)]/|\min_j[u_j^1(e_j) - u_j^0(e_j)]|)$, $j = 1, 2, L, k, u_j^1(e_j) - u_j^0(e_j) \neq 0$ 。

In addition:

(1) $dscc \in [-1, 1]$, the degree of system synergy is positively correlated with its calculation results.

(2) The function of parameter 0 is to ensure the positive and negative effects of system synergy.

3.2.6. *Integration Algorithm of Order Parameters in Supply Chain Synergy Evaluation Model.* The order parameter is mainly calculated by the expert scoring method. Therefore, to avoid the effects on the subjective factors of the rater, the method of expert group judgment can be used to score, and then the expert group decision matrix is constructed to obtain the order parameter component e_{ji} and the order parameter e_j of each subsystem. The specific integration method is as follows [19].

Constructing expert individual decision matrix.

The expert group is composed of P experts. According to the scoring results, the expert set $E = \{E_1, E_2, \ldots, E_P\}$ is obtained, and the individual decision matrix is $M^p = (m_i^p)$, Where: $p = (1, 2, \ldots, p)$, $i = (1, 2, \ldots, n)$.

Calculate the similarity between individual decisions so that the similarity g between individual decisions is as follows:

Let the similarity g between individual decisions be

$$G_i(m_i^p, m_i^q) = \frac{\min\{m_i^p, m_i^q\}}{\max\{m_i^p, m_i^q\}}. \tag{11}$$

Among them, bottle, $p,q = (1,2,\ldots,p)$, $i = (1,2,\ldots,n)$. Table $G_i(m_i^p, m_i^q)$ shows the similarity of decision-making between P and Q experts on the ith index.

Calculate the similarity between individual decision and group decision

$$G_i^p = \frac{1}{p}\sum_{q=1}^{p} G_i(m_i^p, m_i^q). \tag{12}$$

$q = (1, 2, ..., P)$ indicates the similarity between the decision of the ith expert and the group decision for the ith index, and the value of G_i^p is positively correlated with the similarity.

Constructing expert group decision matrix.

The linear weighting method is used to construct the expert decision matrix:

$$M = (m_i),$$
$$m_i = \frac{\sum_{p=1}^{p}(G_i^p \times m_i^p)}{\sum_{p=1}^{p} G_i^p}, \tag{13}$$

$i = (1, 2, \ldots, n)$, m_i represents the integration of all expert decisions for the ith index. The order parameter of SJ can be obtained by integrating the order parameter components e_{ji}, i.e., $e_j = (e_{j1}, e_{j2}, \ldots, e_{jn})$.

3.3. *Cross-Border E-Commerce Logistics Collaborative Evaluation Method and Data Collection*

3.3.1. *Evaluation Method.* The hierarchical analysis process (AHP) is a feasible and comprehensive decision-making method. Analytic hierarchy process (AHP) is a practical decision-making method that combines quantitative and qualitative analysis and quantifies qualitative problems. Based on a qualitative and quantitative combination analysis, the process of analytical hierarchy allows you to effectively structure and synthesize the subjective judgments of people. It is widely used in the research of social, economic, psychological, and organizational management and other systems. When using the process of analytical hierarchy to clarify the scope of the problem, to understand the factors contained in it, to determine the presence of subordination between factors, to analyze and evaluate the research objects with multiobjective and multicriteria by establishing analytic hierarchy process structure model, especially for the system objects with difficult quantitative analysis, and to deal with their qualitative research quantitatively, it can effectively solve the shortcomings of qualitative analysis. The main idea of the analytic hierarchy process is to break down one complex issue by identifying that several influencers associated with ownership constitute performance measures, which are grouped according to their dominant relationships and form a hierarchical structure: build a judgment matrix using the method of two comparisons to determine the relative importance of different indicators in a hierarchy; solve a judgment matrix to get the weight vector corresponding to the maximum value and then perform normalized processing as the weight; determine the consistency of the matrix tested, and the weight of each index is obtained after passing the test. By analyzing the elements contained in complex systems and their related relationships, multilevel analysis and processing make the problems hierarchical and organized and then construct an analytic hierarchy process structure model. Hierarchical process analysis can also sort the indicators, which involves single level sorting; that is, when the indicators of the same level and the indicators of the above level are the comparison criteria, the relative importance scale after mutual comparison is made; it also involves the overall ranking of levels, that is, the relative importance scale (also known as ranking weight vector) of all indicators in the same level to the highest-level indicators (overall objectives) [20]. The main steps of the hierarchical analysis process include: (1) constructing the judgment matrix. According to the

Table 2: Scale method.

Value meaning	1–9 scale
Indicator I is equally important as indicator *J*	1
Indicator I is slightly higher than indicator *J*	3
Indicator I is significantly more important than indicator J	5
Indicator I is very important compared to indicator J	7
The original parts, compared with indicator J	9
The importance of index I is between the above two adjacent levels compared with index J	2,4,6,8
Indicator *J* compared to indicator I	Reciprocal of the above numbers

scaling theory (using Saaty 1–9 scaling method, see Table 2 for details), construct the pairwise comparison judgment matrix, i.e., $A = (a_{ij})_{n*n}(i, j = 1, 2, \ldots, n)$; see Table 3 for details.

In the judgment matrix, there are the following relationships:

① For any i, j satisfy $a_{ij} > 0$, where i, $j = 1,2, \ldots, n$

② For any i, j satisfy $a_{ij} = (1/a_{ji})$, where i, $j = 1,2, \ldots, n$

③ For any i, j satisfy $a_{ii} = 1$, where $i, j = 1,2, \ldots, n$

Sum method or root method shall be applied and normalized. According to $A_w = \lambda_{\max} w$, maximum mode and modal vectors are obtained. Taking the summation method as an example, the specific calculation steps are as follows:

① Matrix A is normalized by column:

$$b_{ij} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} \, (i, j = 1, 2, \ldots, n). \tag{14}$$

② Add the judgment matrix by row:

$$W_i = \sum_{j=1}^{i} b_{ij} \, (i, j = 1, 2, \ldots, n). \tag{15}$$

③ The weight vector can be obtained by normalizing the obtained sum vector:

$$\overline{W}_i = \frac{W_i}{\sum_{i=1}^{1} W_i} \, (i = 1, 2, \ldots, n). \tag{16}$$

④ The maximum special values of the matrix are calculated:

$$\lambda_{\max} = \sum_{i=1}^{n} \frac{[A\overline{W}_i]}{n(\overline{W}_i)_i} (i = 1, 2, \ldots, n). \tag{17}$$

Conduct consistency inspection, and complete the following steps:

① Calculate consistency index:

$$C.I. = \frac{(\lambda_{\max} - n)}{n - 1}. \tag{18}$$

② Find the corresponding mean stochastic consistency index R.I. R.I. is associated with the order of the judgment matrix, as a general rule, the greater the order of the judgment matrix, the greater the chance of random deviation of consistency. See Table 4 for the corresponding relationship.

Table 3: Judgment matrix.

Index	A_1	A_2	......	A_n
A_1	a_{11}	a_{12}	......	a_{1n}
A_2	a_{21}	a_{22}	......	a_{2n}
......	......	......	......	......
A_n	a_{n1}	a_{n1}	......	a_{nn}

Table 4: Average random consistency index of order 1 ~ 10.

Matrix order	1	2	3	4	5	6	7	8	9	10
R.I	0	0	0.52	0.89	1.12	1.26	1.36	1.14	1.46	1.49

③ The percent integrity was calculated:

$$C.R. = \frac{C.I.}{R.I.}. \tag{19}$$

④ When C.R. < 0.1, think that the judgment matrix A passes the consistency check; otherwise, there will be no satisfactory consistency; consideration will need to be given to recreating or correcting the judgment matrix A [21].

4. Results and Analysis

4.1. Analysis of the Dynamic Relationships Based on the VAR Model

4.1.1. VAR Model. The above conclusions are mainly according to the statistical analysis and do not provide insight into the dynamic relationships between electricity enterprises and cross-border logistics, while the model can explain the dynamic model. Correlation of variables in an active system: Because there is a Granger causal relationship between LNKJDS, LNTTL, LNKDL, and LNKDR, the four changes to the company are the changes driven by the VAR model. Using EViews9.o software, select VAR Type for Bayesian VAR, which is limited by the number of samples, so the lag interval of endogenous variables is set as "1–1″ to obtain the corresponding VAR (1) model. From this, the VAR (1) model estimation formula of LNKJDS, LNTTL, LNKDL, and LNKDR can be obtained, as shown in formula

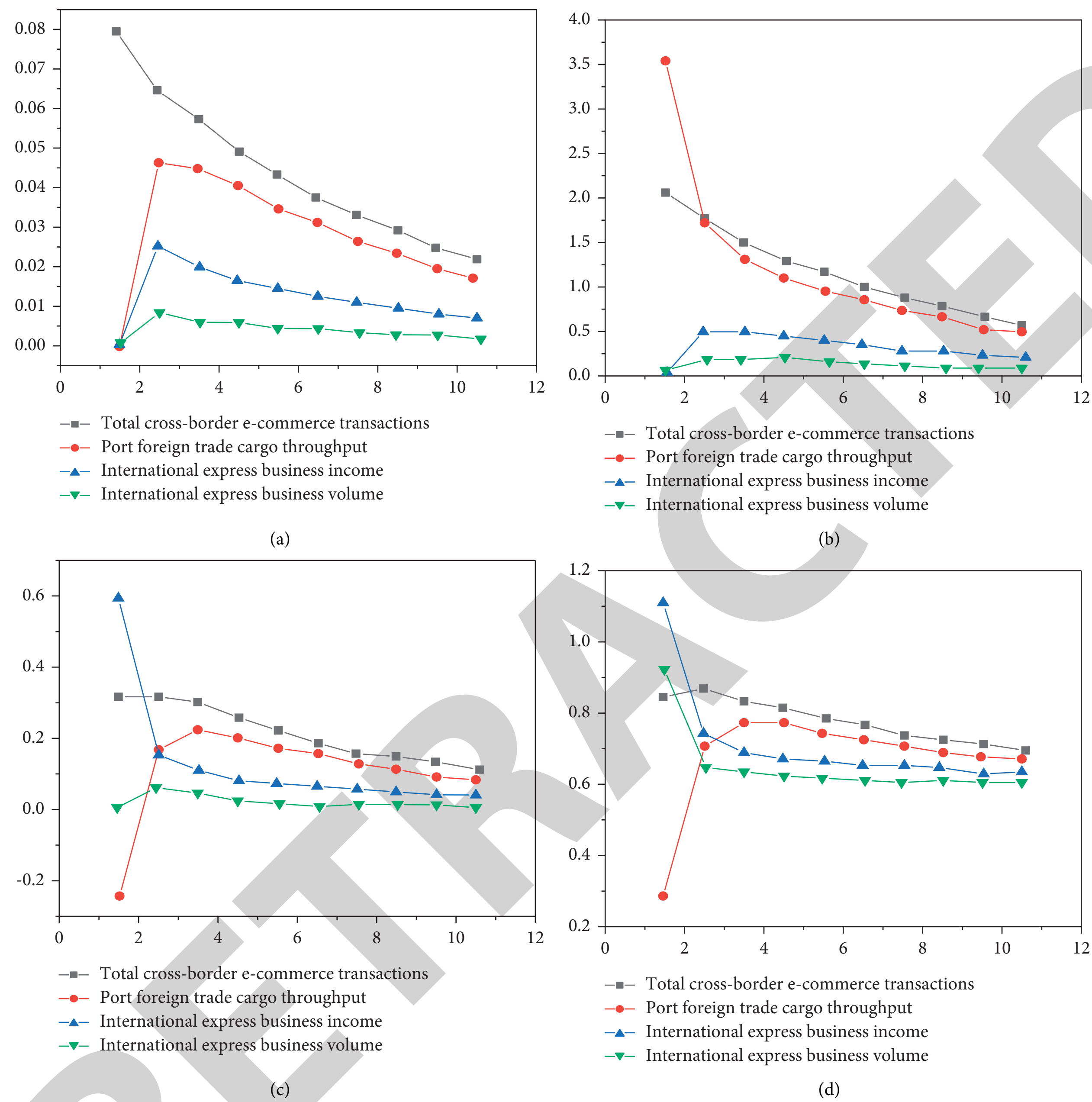


FIGURE 2: Pulse response group plots indicators of cross-border e-commerce and cross-border logistics. (a) Impact of total cross-border e-commerce transactions on each indicator. (b) Impulse response of port foreign trade cargo throughput to various indexes. (c) Impulse response of international express business income to various indicators. (d) Impulse response of international express business volume to various indicators.

$$\begin{pmatrix} LNKJDS_t \\ LNTTL_t \\ LNKDL_t \\ LNKDR_t \end{pmatrix} = \begin{pmatrix} 0.215006 & 1.692483 & 0.132356 & 0.230591 \\ 0.452682 & 0.542035 & 0.025203 & 0.032406 \\ 0.226205 & 1.162573 & 0.142505 & 0.203014 \\ 0.246024 & 0.726504 & 0.068251 & 0.132507 \end{pmatrix} \begin{pmatrix} LNKJDS_{t-1} \\ LNTTL_{t-1} \\ LNKDL_{t-1} \\ LNKDR_{t-1} \end{pmatrix} + \begin{pmatrix} -6.268519 \\ 1.269351 \\ -4.250062 \\ 2.105602 \end{pmatrix}. \tag{20}$$

According to formula (20), it is known that there are cross-border logistics indicators with great impact on the current turnover of cross-border traders, including the most significant impact on the throughput of foreign trade cargo of the country's ports. In all unit roots, the inverse mode is less than 1; that is, it fell in the unit circle, which shows that this research model meets the requirements of the stabilization condition; then you can continue subsequent studies

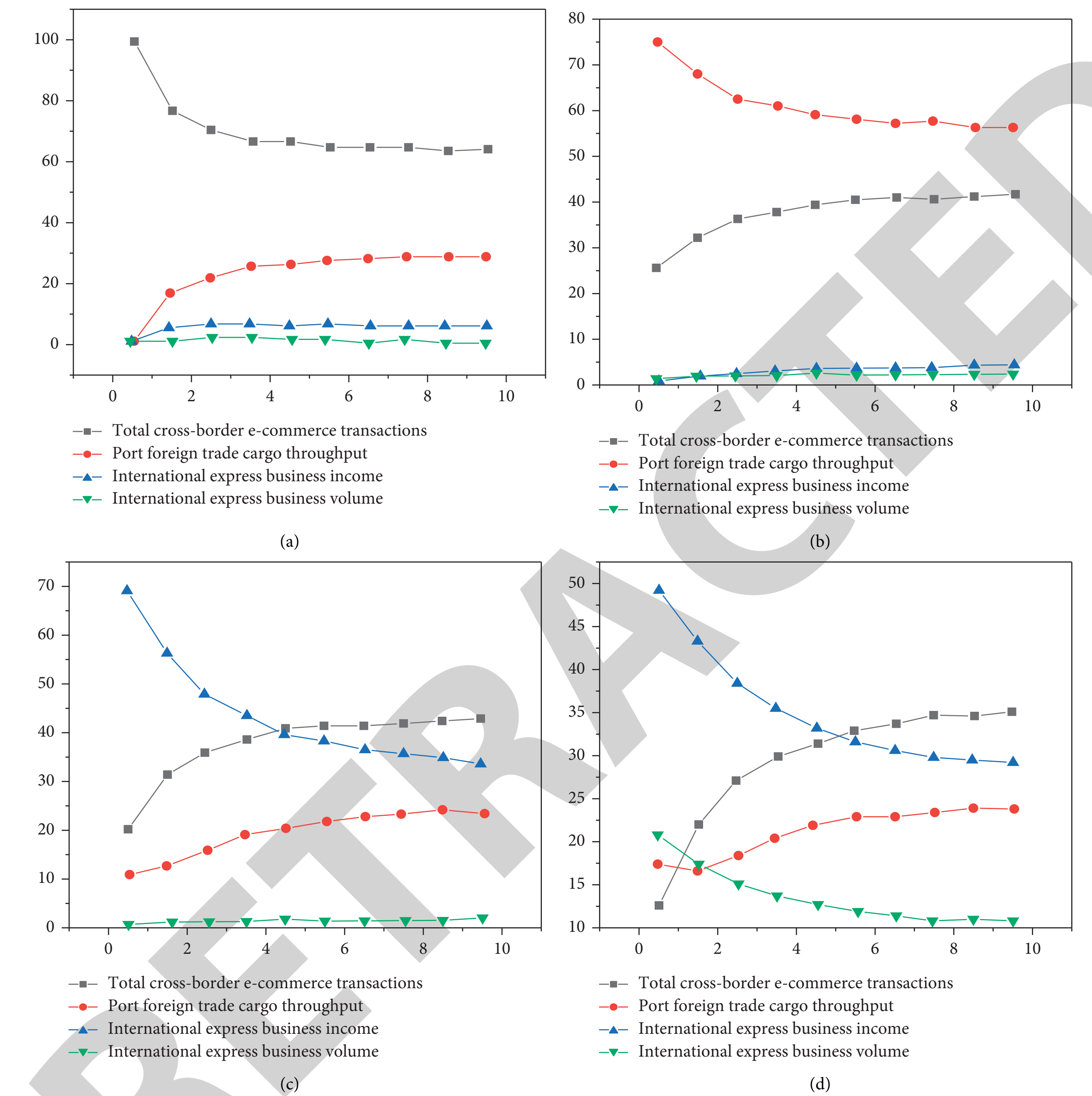


Figure 3: Variance decomposition group diagram of indicators of cross-border e-commerce and cross-border logistics. (a) Variance decomposition of total cross-border e-commerce transactions. (b) Variance decomposition of port foreign trade cargo throughput. (c) Variance decomposition of international express business income. (d) Distribution decomposition international express business volume.

of the impulse response and decomposition of the variance on this basis. This difference can be seen from the comparison in Figure 2.

4.1.2. Variance Decomposition. The pulse response analysis described above shows the effect of the impact of each internal variable on the other internal variables, and the variance decomposition allows further measurement of the degree of interaction between endogenous variables by measuring the contribution of each structure impact to the change. Having done the analysis of the decomposition of the variance, get the histogram of the decomposition of the variance shown in Figure 3 [22].

In combination with the variance analysis table of the VAR (1) model and the variance analysis diagram 3(a), it is known that LNKJDS at the 1st turn only depends on natural fluctuations, gradually weakening from the 2nd turn to about 64%. The level of contribution of the three indicators of cross-border logistics to the dispersion of fluctuations in the growth of the total volume of cross-border transactions with electric enterprises begins to appear from the 2nd

Table 5: Three-level assessment index system.

Level 1 indicators	Secondary indicators	Level 3 indicators
Cross-border e-commerce ecosystem collaboration (U)	Species synergy (U_1)	Impact of core species (U_{11})
		Impact of key species (U_{12})
		Impact of supporting species (U_{13})
		Effects of parasitic species (U_{14})
	Environmental collaboration (U_2)	Influence of political environment (U_{21})
		Impact of economic environment (U_{22})
		Impact of social environment (U_{23})
		Impact of technology environment (U_{24})
	Supply chain collaboration (U_3)	Supplier collaboration between suppliers (U_{31})
		Collaboration between supplier and trading platform (U_{32})
		Collaboration between trading platform and consumers (U_{33})
	Geospatial collaboration (U_4)	Exporting country space collaboration (U_{41})
		International space collaboration (U_{42})
		Input country space collaboration (U_{43})
		Reverse spatial collaboration (U_{44})

edition, including the maximum level of the contribution of LNTTL, which tends to flatten from the beginning of the 2nd edition to about the 4th edition and eventually stabilizes at around 29%; while the level of contribution of LNKDR and LNKDL is relatively small, LNKDR has been stable around 6% since the 2nd turn for a long time, and LNKDL has been trending around 1% for a long time. In fact, in the long term, although it can also explain the volatility of the cross-border logistics industry, with one-third of the changes in the development of the cross-border electrical industry, changes in the development of the cross-border electrical industry are mainly influenced through the volatility of the industry itself. The analysis in charts 3(b)–3(d) gives an idea of how error volume of cross-border commercial transactions contributes to changes in the performance of cross-border logistics. In the long term, the contribution of gross profit volume of cross-border commercial transactions to the change in thread indicator in intersection edge logging, respectively, is about 41% (LNTTL), 42% (LNKDR), and 35% (LNKDL) [23]. In general, the degree of contribution of fluctuations in the total volume of cross-border commercial transactions to the change in various indicators of cross-border logistics is almost 40 percent; cross-border development electrical industry has a stronger southern impact of cross-border logistics.

4.2. Fuzzy Comprehensive Evaluation

4.2.1. Determine the Evaluation Object. Cross-border e-commerce ecosystem collaboration is determined as the overall evaluation objective and set as the evaluation object of fuzzy comprehensive rating, that is, u = ”cross-border e-commerce ecosystem collaboration”.

4.2.2. Establish Evaluation Subobjective Set. Combined with the above research results, the four levels affecting cross-border e-commerce ecosystem collaboration are species collaboration, environmental collaboration, supply chain collaboration, and geospatial collaboration, and then the evaluation subgoal set is constructed. Set U_1 = ”species collaboration”, U_2 = ”environment collaboration”, U_3 = ”supply chain collaboration”, and U_4 = ”geospatial collaboration” to evaluate the subtarget set $U = (U_1, U_2, U_3, U_4)$.

4.2.3. Construction of Evaluation Index System. By subdividing and analyzing the influencing factors of the subtarget set, that is, each subtarget U_i is affected by each index $u_{i1}, u_{i2}, \ldots, u_{it}$, so as to construct the index set u_{ij}, set $u_{ij} = (u_{i1}, u_{i2}, \ldots, u_{it})$, where $i = 1, 2, \ldots, s$. Combined with the analytic hierarchy process, the evaluation object is set as the first level index, the evaluation subobjective set is set as the second level index, and the index set is set as the third level index, so as to construct the three-level evaluation index system of cross-border e-commerce ecosystem coordination, as shown in Table 5. At the level of species synergy, the influencing factors are determined as four three-level evaluation indexes: the influence of core species, the influence of key species, the influence of supporting species, and the influence of parasitic species. At the level of environmental synergy, four three-level evaluation indicators are determined: the impact of political environment, the impact of economic environment, the impact of social environment, and the impact of technological environment. At the level of supply chain collaboration, three three-level indicators are determined: supplier collaboration between suppliers, supplier collaboration with trading platform, and trading platform collaboration with consumers. At the geospatial level, four three-level indicators are determined: spatial synergy of exporting countries, international spatial synergy, spatial synergy of importing countries, and reverse spatial synergy [24].

4.2.4. System for Determining Evaluation Index Weight. From the cross-border e-commerce Eagle bear exchange members, three experts were selected to form an expert opinion group. Table 6, Table 7, and Table 8 are obtained

Table 6: Evaluation scores of expert 1.

Index	U_1	U_2	U_3	U_4
U_1	1	4	2	3
U_2	1/4	1	1/3	1/2
U_3	1/2	3	1	3
U_4	1/3	2	1/3	1

Table 7: Evaluation scores of experts 2.

Index	U_1	U_2	U_3	U_4
U_1	1	3	2	2
U_2	1/3	1	1/2	1/2
U_3	1/2	2	1	2
U_4	1/2	2	1/2	1

Table 8: Evaluation scores of experts 3.

Index	U_1	U_2	U_3	U_4
U_1	1	3	2	4
U_2	1/3	1	1/4	1/3
U_3	1/2	4	1	2
U_4	1/4	3	1/2	1

Table 9: Judgment matrix of average value of expert opinions.

Index	U_1	U_2	U_3	U_4
U_1	1.0000	3.3333	2.0000	3.0000
U_2	0.3056	1.0000	0.3611	0.4444
U_3	0.5000	3.0000	1.0000	2.3333
U_4	0.3611	2.3333	0.4444	1.0000

according to the evaluation scores of experts on the secondary index $U_i\,(i = 1, 2, \ldots, 4)$.

Take the average of the evaluation results of the expert opinion group and get Table 9.

5. Conclusion

The crossover business of the power sector is currently underway in the early stage of development, and although much attention is being paid to it, the available research results are still limited and there is a lack of systematic, mature research result and rational, mature theoretical research support. Electric dealers and logistics are in symbiosis with each other, interact with each other, and facilitate each other, and these relationships equally exist in cross-border e-commerce activities. Synergistic research in the field of cross-border e-commerce, especially cross-border traders and cross-border, is of great importance and already imminent. In this context, combing the development trends of cross-border merchants and cross-border logistics, relying on relevant research results at home and abroad, using the value chain, synergy theory, and ecosystem theory as a research city, evaluating the current state of cross-border e-commerce ecosystem synergy, from which the key elements that affect the synergy of cross-border e-commerce are extracted, and studying the problems of synergy between cross-border merchants and cross-border logistics, the following conclusion can be drawn: from the point of view of the ecosystem of cross-border e-commerce, one of the central issues is the synergy of cross-border logistics of merchants. Using the theory of the structural equation model, a model of the joint theory of cross-border logistics of the electric enterprise is proposed. Based on the theoretical model, the corresponding research assumptions are proposed. As a result of multiple test trials, the cross-border electric enterprise logistics joint scale was determined, and the cross-border electric enterprise logistics joint structure equation model was built to develop the cross-border electric enterprise logistics joint verification questionnaire for the blue book. In combination with the data obtained from the questionnaire, using the statistical analysis software SPSS, AMOS, it was found that the data and the structural equation model are of high suitability and can be applied to analyze the results of the study. The final verification results show that cross-border e-commerce collaboration with other species has a significant positive impact on cross-border e-commerce logistics collaboration, while cross-border e-commerce collaboration with other species has no significant positive impact on cross-border logistics chain collaboration. However, synergy between the cross-border electricity industry and other species does not have a significant positive impact on the synergy of crossover supply chains.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This work was supported by the Hebei Jiaotong Vocational and Technical College.

References

[1] C. Di, "Research on the product logistics cost control strategy based on the multi-source supply chain theory," *Intelligent Automation & Soft Computing*, vol. 26, no. 3, pp. 557–567, 2020.

[2] I. Ghalehkhondabi and R. Maihami, "Sustainable municipal solid waste disposal supply chain analysis under price-sensitive demand: a game theory approach," *Waste Management & Research*, vol. 38, no. 3, pp. 300–311, 2019.

[3] K. Matsui, "Optimal bargaining timing of a wholesale price for a manufacturer with a retailer in a dual-channel supply chain," *European Journal of Operational Research*, vol. 287, no. 1, pp. 225–236, 2020.

[4] S. Guan, "Smart e-commerce logistics construction model based on big data analytics," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 1–9, 2020.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9838165, 1 page
https://doi.org/10.1155/2023/9838165

Retraction

Retracted: Analysis of International Competitiveness of China's Mobile Phone Industry Based on Data Mining Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Lu and Z. Lv, "Analysis of International Competitiveness of China's Mobile Phone Industry Based on Data Mining Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 8460574, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8460574, 10 pages
https://doi.org/10.1155/2022/8460574

Research Article

Analysis of International Competitiveness of China's Mobile Phone Industry Based on Data Mining Algorithm

Jun Lu and ZhongLi Lv

School of Economics and Management, Guizhou Normal University, Guiyang 550001, China

Correspondence should be addressed to Jun Lu; lujun234@gznu.edu.cn

Received 17 February 2022; Revised 16 March 2022; Accepted 21 March 2022; Published 23 April 2022

Academic Editor: Chin-Ling Chen

In order to analyze the international competitiveness of China-made mobile phones, this paper combines data mining algorithms to analyze the international competitiveness of China's mobile phone industry, improve the international market share of China-made mobile phones, and study the concept of multiattribute problems and related research methods. Moreover, the paper carries out the model description of the selected research methods combining principal component analysis, data envelopment analysis, and sorting method approaching the ideal solution. In addition, this paper expounds on the principles and models of PCA, DEA, and TOPSIS and selects an intelligent algorithm suitable for this model. Finally, this paper verifies the validity of the model proposed in this paper, conducts statistical analysis through the data mining model, evaluates the data mining effect through multiple simulation exercises, and verifies the validity of the system model.

1. Introduction

Faced with the problems of high market share and low profit margins of domestic mobile phone companies, clustering of low-end and middle-end markets, weakness in the high-end market, and low profits, in the fierce market competition, timely adjustment of business strategies to enhance the competitiveness of domestic mobile phone companies has become a top priority. At the same time, as consumers of corporate products and services, the perceived value of customers can be said to be a new and important source of corporate competitiveness. Therefore, it is necessary to proceed from the perspective of customer value. If customers have high perceived value for a certain mobile phone product, then the company has good competitiveness. In order to gain strong competitiveness, domestic mobile phone companies should establish a customer-oriented business philosophy, enhance customer perceived value, and regularly monitor and utilize customer perceived value, so as to provide customers with mobile phone products and services with higher customer value. Therefore, thinking about how to improve the competitiveness of domestic mobile phone companies from the perspective of customer value is an important research idea.

This article combines data mining algorithms to analyze the international competitiveness of China's mobile phone industry, increase the market share of domestic mobile phones in the world, and provide a reference for the subsequent development of our country's mobile phone industry.

2. Related Work

The theory of enterprise competitiveness originates from the research on the acquisition and maintenance of competitiveness. According to the origin of enterprise competitiveness, it can be divided into two major theoretical schools: exogenous theory and endogenous theory [1]. The school of exogenous theory focuses on the analysis of the external competitive environment of enterprises and believes that the competitiveness of enterprises is reflected in the comparison with competitors. Literature [2] points out that enterprises design, manufacture, and sell goods or provide services in their own environment. When compared with other competitors, the ability and opportunity that is more attractive in price and quality is the competitiveness of enterprises; literature [3] believes that compared with competitors, the ability of enterprises to acquire, create, and use knowledge is stronger or weaker. It can be characterized as the competitiveness of enterprises. Michael Porter attributed the dominant position of enterprises in the industry to the competitiveness of enterprises. Literature [4] pointed out that the core of enterprise competitiveness is comparative productivity, which itself is based on the interindividual, a relative concept obtained by comparison. When the above scholars define the competitiveness of enterprises, although they consider different angles, they all focus on the analysis of the external competitive environment of enterprises, which belongs to the school of exogenous theory. The endogenous school focuses on the analysis of the internal situation of the enterprise. Literature [5] believes that the human and material resources of the enterprise itself form the competitiveness of the sustainable and long-term development of the enterprise. Literature [6] attributes the competitiveness of the enterprise to the interior of the enterprise. Domestic scholars also hold similar views. Literature [7] believes that enterprise competitiveness is the ability to face the market and customers.

In the exploration of the connotation of technological competitiveness of enterprises from different perspectives, the views of various scholars have not been unified, but they can be basically divided into two schools: the ability school and the resource school. With the deepening of research studies, the literature [8] integrated the viewpoints of the capability school and the resource school, believed that the enterprise itself is a combination of resources and capabilities, and pointed out that the technological competitiveness of enterprises stems from the full use of internal and external technological innovation resources, especially the full use of internal and external technological innovation resources. The role of the unique, scarce, and irreplaceable technological assets owned by an enterprise enables itself to provide more attractive products or services in the market and to obtain more long-term benefits than its competitors. By evaluating the technological competitiveness of an enterprise, it can promote the efficient and rational use of all its own resources and maximize the ability of the enterprise. With the passage of time, the importance of technology in the development of enterprises, especially high-tech enterprises, has become increasingly significant. If an enterprise wants to surpass or defeat its competitors, it must improve its technological competitiveness [9].

Literature [10] comprehensively uses the analytic hierarchy process, the entropy weight coefficient method, the Delphi method, and the regression analysis method to evaluate the technological competitiveness of our country's self-owned brand automobile enterprises with the technological input and output capabilities of enterprises as the breakthrough point. Literature [11] verified the uniqueness and applicability of the new evaluation theory of cross analysis and the corresponding cross-analysis method for the evaluation of enterprise technological competitiveness through empirical research and provided new research ideas and research methods for the study of enterprise technological competitiveness. Literature [12] refines the technological competitiveness of enterprises according to the three main stages of technology research and development, technology integration, and technology monopoly in the process of independent innovation and establishes a structural equation model based on the path relationship between strategic orientation, technological competitiveness, and enterprise status. Literature [13] designed the evaluation index of IoT technology competitiveness based on the "push-pull model" and "human technology symbiosis model." Literature [14] uses the AHP, based on the patent-based LED enterprise technology competitiveness evaluation system, analyzes the innovation input and output data of LED packaging listed enterprises, and evaluates the enterprise's technological competitiveness. Literature [15] uses the relative technical advantages of patents to define the technical strength of each research object and draws a patent portfolio chart with the technical attractiveness of each company's mainstream technology, the relative position of patents, and technical strength as indicators. Literature [16] emphasizes that brands have both functional value and emotional value.

3. Data Mining Algorithm Based on Principal Component Analysis

In data analysis, we often face the problems of judging the pros and cons of things and mutual rules. However, the factors that affect the characteristics of a certain thing and its development law are diversified. In order to analyze more deeply, we need to analyze and evaluate various influencing factors related to it. However, multivariate and large sample data will bring about problems such as multicollinearity, and the duplication of information reflected by each influencing factor will affect the authenticity and scientificity of the statistical results. Therefore, in order to avoid information overlap and reduce workload as much as possible, people put forward the idea of "dimensionality reduction." Principal component analysis is performed by finding a few uncorrelated variables, calculating their linear combination, and transforming them into a few comprehensive indicators. At the same time, it saves most of the information in the original variable data, which is the most widely used multivariate statistical analysis method.

The principal components are several comprehensive indexes formed by the original indexes through a series of mathematical calculations, that is, F_1, F_2, and F_3 in the above example. According to the amount of information contained in the principal components, they are called "first principal component," "second principal component," "third principal component," and so on.

3.1. Mathematical Model of Principal Components. $X_1, X_2, X_3, \ldots, X_p$ is a p-dimensional random variable, and principal component analysis is done to transform p

observed variables into p new indicators through linear combination, namely,

$$\begin{aligned}
F_1 &= \mu_{11}X_1 + \mu_{12}X_2 + \cdots + \mu_{1p}X_p,\\
F_2 &= \mu_{21}X_1 + \mu_{22}X_2 + \cdots + \mu_{2p}X_p,\\
M&,\\
F_p &= \mu_{p1}X_1 + \mu_{p2}X_2 + \cdots + \mu_{pp}X_p.
\end{aligned} \tag{1}$$

The model meets the following conditions:

(1) $\mu_{i1}^2 + \mu_{i2}^2 + \cdots + \mu_{ip}^2 = 1$

(2) $\mathrm{Cov}(F_i, F_j) = 0, \quad i \neq j, i, \ j = 1, 2, \ldots, p$

(3) $\mathrm{Var}(F_1) \gg \mathrm{Var}(F_2) \gg \gg \mathrm{Var}(F_p)$

The purpose of principal component analysis is to simplify variables, and the number of principal components is usually less than the number of original variables. However, as for the actual problems, several principal components should be retained. We naturally hope that the principal components reflect as much information of the original variables as possible. Therefore, it is necessary to weigh the number of principal components and the retained information. The "information" here is measured by variance; that is, the larger $\mathrm{Var}(F_1)$ is, the more information F_1 contains. In the end, the new indicators $F_1, F_2, F_3, \ldots F_k (k \leq p)$ fully retain the main information in the original variables and are independent of each other.

3.2. Principal Component Analysis. The specific steps of principal component analysis are as follows:

(1) The algorithm calculates the correlation coefficient matrix.

(2) The algorithm calculates the characteristic root of the correlation coefficient matrix and the corresponding characteristic vector.

(3) The algorithm selects the largest feature root, and the corresponding feature vector is equal to the coefficient of the first principal component. Moreover, the algorithm selects the second largest feature root, and the corresponding feature vector is equal to the coefficient of the second principal component.

(4) The algorithm calculates the cumulative contribution rate and selects the appropriate number of principal components.

(5) The algorithm writes the expressions of the first k principal components.

Data envelopment analysis is widely used in operations research, management science, and economics. In the production activities of enterprises, due to limited resources, producers always want to get the most output with the least input. From the perspective of input and output, the DEA method is a new input-output dual-criteria model, which minimizes input while ensuring maximum output. The DEA method is a nonparametric estimation method, which is suitable for dealing with problems that require decision-making or evaluation indicators, and the goal itself does not have to have a clear functional form. Specifically, the DEA method, based on linear programming theory, performs a series of mathematical calculations on input index and output index data and is a quantitative analysis method that evaluates the relative effectiveness of decision-making cases with the same attributes.

The DEA method mainly evaluates the relative efficiency of the decision-making unit (DMU). Each DMU has the same input variables and output variables. The DEA method obtains the comprehensive efficiency index of each DMU by calculating the weighted ratio of output and input data and ranks the DMUs according to this value to determine the effective DMU, which is the decision-making unit with the highest relative efficiency.

Next, we will introduce the classic DEA model: the C^2R model. We assume that there are n decision-making units and each decision-making unit has m kinds of input variables and s kinds of output variables, which, respectively, represent the "resources consumed" and "effects of work" of the decision-making unit, in Figure 1.

Now, we want to evaluate the efficiency of the i_0-th decision-making unit. For the convenience of operation, DMU_i is abbreviated as DMU_0, (X_i, Y_i) as (X_0, Y_0), and h_6 as h_0. The classic C^2R model optimizes the weight coefficients u and v to maximize h_0, which is used to solve the following optimization problem:

$$\begin{aligned}
&\max h_0 = \frac{\sum_{r=1}^{s} \mu_r y_{r0}}{\sum_{j=1}^{n} v_j x_{j0}},\\
&\frac{\sum_{r=1}^{s} \mu_r y_{ri}}{\sum_{j=1}^{n} v_j x_{ji}} \leq 1, \quad i = 1, \ldots, n,\\
&u_r > 0, \quad r = 1, \ldots, s,\\
&v_j > 0, \quad j = 1, \ldots, m.
\end{aligned} \tag{2}$$

Here, y_{ri} is the output of the i-th decision-making unit for the r-th type of output indicator and x_{ji} is the i-th decision-making unit's input for the j-th type of input indicator. u_r and v_j are the weight coefficients corresponding to the optimal solution of the above-mentioned maximum problem.

In the above-mentioned optimization problem, for each h_0^*, we can obtain a set of optimal μ_r^* and v_j^*. If the maximum efficiency ratio among them reaches 1, then we call the decision unit to be technically effective. A decision-making unit that is technically effective means that it does not need to increase any output or reduce any input. The closer the relative efficiency value is to 1, the higher the energy consumption efficiency of the decision-making unit.

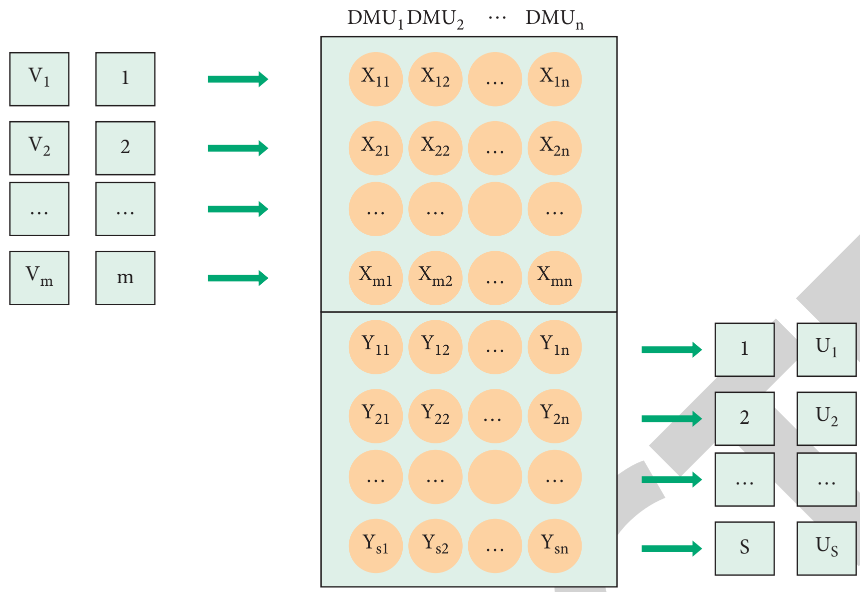


Figure 1: Input and output of the decision-making unit.

We set as follows:

$$\begin{aligned} t &= \frac{1}{\sum_{j=1}^{m} v_j x_{j0}}, \\ \omega &= t v, \\ \mu &= t \mu. \end{aligned} \tag{3}$$

Then, the original fractional planning can be transformed into

$$\begin{aligned} &\max \sum_{r=1}^{s} \mu_r y_{r0}, \\ &\sum_{j=1}^{m} \omega_j x_{ji} - \sum_{r=1}^{s} \mu_r y_{ri} \geq 0 \quad , i = 1, 2, \ldots, n, \\ &\sum_{j=1}^{m} \omega_j x_{j0} = 1, \\ &\mu_r \geq 0, \quad r = 1, 2, \ldots, s, \\ &\omega_j \geq 0, \quad j = 1, 2, \ldots, m. \end{aligned} \tag{4}$$

The dual problem of linear programming is as follows:

$$\begin{aligned} &\min \theta, \\ &\sum_{i=1}^{n} X_i \lambda_i + s^- = \theta X_0, \\ &\lambda_i \geq 0; \quad i = 1, 2, \ldots, n; \ s^+ \geq 0, \ s^- \geq 0. \end{aligned} \tag{5}$$

Here, s^+ and s^- are the slack variables.

The above is the mathematical model of the DEA method. The biggest advantage of this method is it being simple and easy to understand. In practical applications, the algorithm is also relatively easy to implement. It can be solved by special DEAP2.0 software or by writing a Matlab program. However, the DEA method also has some shortcomings in solving practical problems. We usually require that the number of decision-making units should not be less than twice the output index, and it is also necessary to ensure that the data are nonnegative. In addition, if there is a large correlation between input variables and output variables, DEA's evaluation results will be affected. This can be solved ingeniously by combining this with PCA. Therefore, before using DEA for efficiency evaluation, this paper first uses PCA to reduce the dimensionality of the indicators and replaces all the original indicators with a few principal components.

In fact, the DEA method only obtains the calculation results of relative efficiency, that is, determines which decision-making units are relatively effective or uses resources relatively efficiently, but cannot obtain the detailed characteristics and properties of the data of these effective units and other information. That is to say, just calculating the utilization efficiency of resource input and output using decision-making units cannot fundamentally explain the reasons for selecting certain decision-making unit gap between alternatives. In order to solve this problem, this paper further analyzes the results calculated by the DEA method through the TOPSIS method and studies the preference value of the data contained in all decision-making units, so as to make a more comprehensive assessment of the comprehensive competitiveness of each cruise home port.

The sorting method approximates the ideal solution (technique for preference by similarity to the ideal solution, abbreviated as TOPSIS). The preference value can be converted into Euclidean distance for calculation and measurement, and after comprehensive comparison, an optimal plan and the comparison of all plans can be obtained.

The basic idea of the TOPSIS method is to consider the distance between the candidate scheme and the most ideal scheme and the least ideal scheme at the same time. However, there is usually no such positive ideal solution and negative ideal solution in the original scheme set. Therefore, if there is a solution in the scheme that is both closest to the positive ideal solution and farthest from the negative ideal solution, this solution is the best solution in the scheme set plan; otherwise, it is the worst plan. When applying the TOPSIS method, it is necessary to determine the weight of an evaluation index. Scholars at home and abroad are actively studying scientific and reasonable weight value determination methods to solve practical problems, and now, they have achieved good results.

The calculation steps of the TOPSIS method are described as follows:

(1) First, the algorithm converts the original data into a decision matrix. The decision matrix is defined as follows:

$$C = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}. \tag{6}$$

Here, m is the number of alternatives, n is the number of evaluation indexes, and x_{ij} is the evaluation value of the j-th index of the i-th program.

(2) Since the dimensions of the selected evaluation indicators are different, in order to facilitate comparison, we need to standardize the following initial indicator data: $R = (r_{ij})_{m\times n}$, where $r_{ij} = x_{ij}/\sum_i x_{ij}^2,\quad i = 1,\ldots,m;\ j = 1,\ldots,n.$

(3) The algorithm can get the weight value of each evaluation index (attribute) $W_j\,(j = 1,\ldots,n)$. Next, the algorithm defines the weighted standardized decision matrix as V, and the calculation of $V = R\times W$ is as follows:

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix}. \tag{7}$$

Here, $v_{ij} = r_{ij}\times w_j$ and $\sum_{j=1}^{n} w_j = 1$.

(4) The algorithm determines the positive ideal solution A^+ and the negative ideal solution A^-.

$$\begin{aligned}
&\text{Positive ideal solution: } A^+ = \left(v_1^+, v_2^+, \ldots, v_j^+, \ldots, v_n^+\right),\\
&\text{negative ideal solution: } A^- = \left(v_1^-, v_2^-, \ldots, v_j^-, \ldots, v_n^-\right),\\
&\text{where } v_j^+ = \begin{cases} \max\left(v_{ij}\right), j\in J^+ \\ \min\left(v_{ij}\right), j\in J^- \end{cases} (j = 1,2,\ldots,n),\\
&v_j^- = \begin{cases} \max\left(v_{ij}\right), & j\in J^- \\ \min\left(v_{ij}\right), & j\in J^+ \end{cases} (j = 1,2,\ldots,n)
\end{aligned} \tag{8}$$

J^+ is the benefit-type attribute collection and J^- is the cost attribute collection.

(5) The algorithm calculates the distance. Here, we use the Euclidean distance to calculate the distance from each alternative to the positive ideal solution A^+ and the negative ideal solution A^-, respectively.

$$\begin{aligned}
S_i^+ &= \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^+\right)^2} \quad (i = 1,2,\ldots,m),\\
S_i^- &= \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^-\right)^2} \quad (i = 1,2,\ldots,m).
\end{aligned} \tag{9}$$

The algorithm calculates the relative closeness to the positive ideal solution A^+ and the relative closeness is as follows:

$$C_i = \frac{S_i^-}{\left(S_i^+ + S_i^-\right)}. \tag{10}$$

(6) The algorithm arranges the preference order, sorts the alternatives according to the descending order of the value of C_i, and chooses the one with the largest value of C_i as the optimal scheme.

After the above steps, we can apply the TOPSIS method to select the best scheme from a large number of alternative schemes and, at the same time, sort all the schemes by preference value.

4. Analysis of International Competitiveness of China's Mobile Phone Industry Based on Data Mining Algorithms

Figure 2 intuitively shows the relationship among the competitiveness of mobile phones, the competitiveness of the mobile phone industry, and the competitiveness of mobile phone companies.

Figure 2 shows that corporate competitiveness, industrial competitiveness, and national competitiveness are in a progressive relationship, and the former cannot be simply summed up to get the latter. The relationship between the three is like a pyramid, and the bottom-up shows a clustering effect. In other words, the advantages of the lower layer can be summarized and refined to the upper layer to influence the upper layer. From the top to bottom, it shows a penetration effect; that is, the policies of the upper level will act on the lower level.

The determinants of the competitive advantage of a country, industry, or enterprise can be summarized as four basic determinants and two auxiliary factors. The four basic determinants are production factors, demand conditions, related and supporting industries, and corporate strategy, structure, and competition in the same industry. The two supporting factors are government and opportunity. Each factor has far-reaching significance, and different factors are related to each other, as shown in Figure 3.

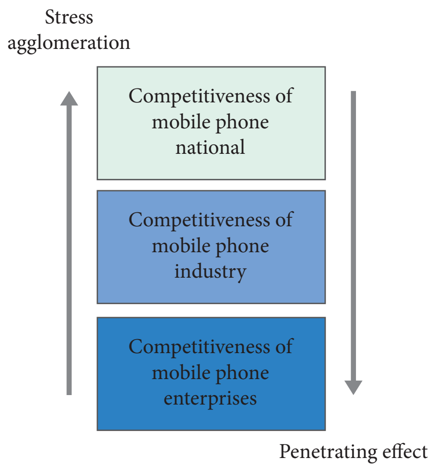


Figure 2: The relationship between national competitiveness, industrial competitiveness, and enterprise competitiveness.

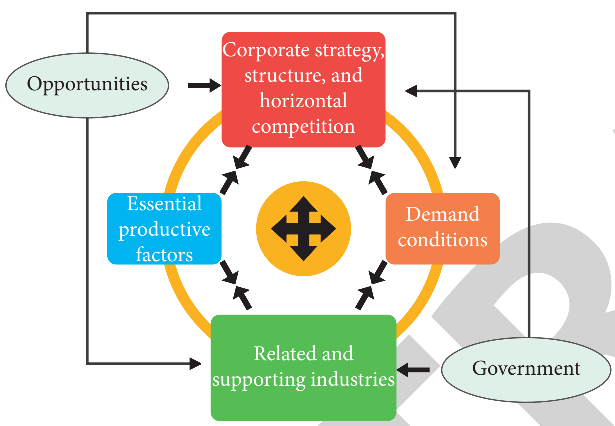


Figure 3: Diamond model.

The three perspectives of environmental competitiveness, enterprise competitiveness, and market competitiveness are used to evaluate the international competitiveness of the industry. Environmental competitiveness is the basis for the competitiveness of an industry, including the country's industrial policy and industrial layout, domestic demand conditions, and foreign direct investment. Whether these factors have advantages directly determines whether the industry has international competitiveness. The competitiveness of an enterprise is the core of whether an industry is competitive, including the economies of scale of the enterprise, the labor productivity of its employees, and the R&D capabilities. As a microsubject, an enterprise can rationally use resource endowments and realize the optimal allocation of resources, which is a decisive condition for an enterprise to have a competitive advantage. Market competitiveness is a supplementary condition for whether an industry is competitive, including revealed comparative advantage index (RCA), international market share, and trade competitiveness index. These factors can help the mobile phone industry to expand its competitiveness. Similarly, the lack of market competitiveness will directly lead to the low competitiveness of the industry, as shown in Figure 4.

After years of development, China's mobile phone industry has formed a multidimensional and interactive industrial chain. Due to the different expression methods of different experts, based on the opinions of other experts and scholars, this article summarizes and describes China's mobile phone industry chain as shown in Figure 5:

The activities of the value chain are independent of each other, but they are also connected. It classifies the value activities of enterprises according to assets, personnel, functions, and values. It is necessary to well identify the basic activities and auxiliary activities, production and operation, sales, and services. Sometimes, a basic value chain can also be decomposed again, so the various activities of the enterprise must be strictly distinguished, neither overlapping nor missing. The value chain of a mobile phone company can be expressed as shown in Figure 6.

The life cycle characteristics of mobile phone products can be represented by the change trend of the product value in the R&D phase, product investment period, growth period, maturity period, and decline period, which can be further illustrated in Figure 7. In the figure, the AB stage represents the initial development stage of the product. It can be seen that the value generated by the product at this stage is negative, the company is in the preassessment stage of the product, and there is only input and no corresponding output. BC represents the input period of the product. It has a small batch of samples output from the acquisition of the hard mold, so its curve shows an upward trend, which is close to the X axis. CD represents the growth period of the product. During this period, the demand for products has increased rapidly and it has begun to enter the stage of mass production. Since the products in this period just entered the market, the price of the products was relatively high, so the value of output increased significantly and showed a close-to-straight upward trend.

Combined with the above model analysis, this article combines the third part of the algorithm to build a data mining system. Data mining is a process of extracting valuable information and data from a large amount of data. A large amount of historical data is generated in various operation links, and these data are an intangible asset of the enterprise. Which company can effectively and deeply mine the value of data can better provide customers with efficient products or services. Because different industries use different analytical techniques and tools, the analytical methods can be unified and coordinated to solve many valuable business problems. The data mining model proposed in this paper is shown in Figure 8.

On the basis of the above research, the effectiveness of the model in this paper is verified, the data mining model is used for statistical analysis, the data mining effect is evaluated through multiple simulation exercises, and the results shown in Figure 9 are obtained.

This paper evaluates the clustering effect of the system constructed in this paper in the analysis of the international competitiveness of the mobile phone industry, and the results shown in Figure 10 are obtained.

Through cluster analysis, the industrial chain should be integrated to enhance the comprehensive competitiveness of

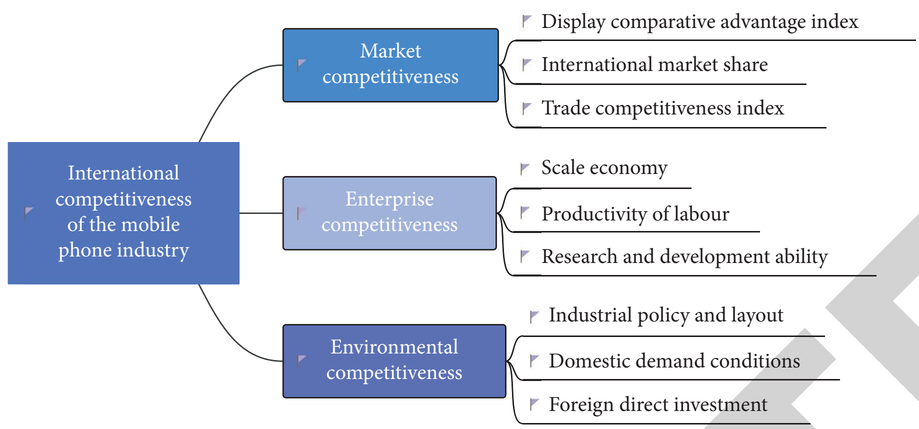


FIGURE 4: Evaluation method of the international competitiveness of the mobile phone industry.

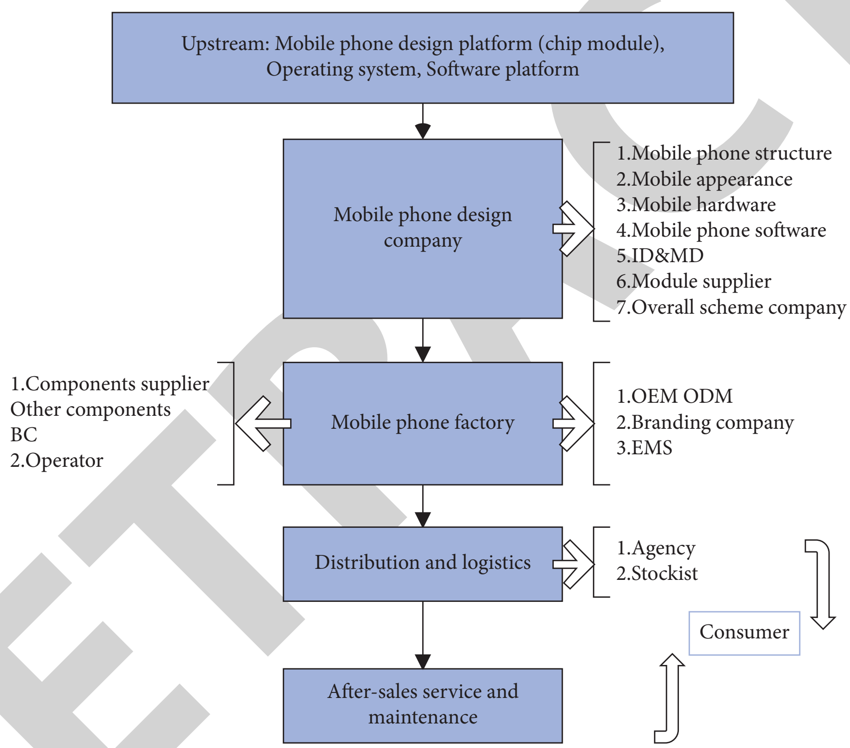


FIGURE 5: Schematic diagram of China's mobile phone industry chain.

local enterprises. It is necessary to strengthen cooperation between domestic mobile phone companies, achieve group breakthroughs, carry out industrial upgrades, and jointly promote the healthy development of domestic mobile communication terminal products with the power of groups. Moreover, it is necessary to strengthen the cooperation between domestic mobile phone companies and mainstream mobile phone design companies to make up for the inherent shortcomings of domestic mobile phone back-ends (independent research and development of core technologies, etc.), so that R&D products can maintain a close tracking advantage with international brands. At the same time, it is necessary to strengthen the strategic cooperation between domestic mobile phone companies and operators to reduce the risks of mobile phone production and to diversify business projects with the help of various value-added services of operators, expand the market, and seek new profit growth points. It is necessary to systematically strengthen brand building, clarify product positioning, enhance brand value, and avoid excessive price competition under homogenization. It is necessary to deeply tap the potential of the third and fourth tier markets, complete the accumulation of enterprise strength, prepare for the first and second markets, actively explore foreign markets, highlight the encirclement, and reverse the situation with exports. It is necessary to strengthen the internal management of domestic mobile phone companies to ensure the implementation of the system. In terms of enhancing core competitiveness, it is necessary to increase technological research and development, introduce advanced design

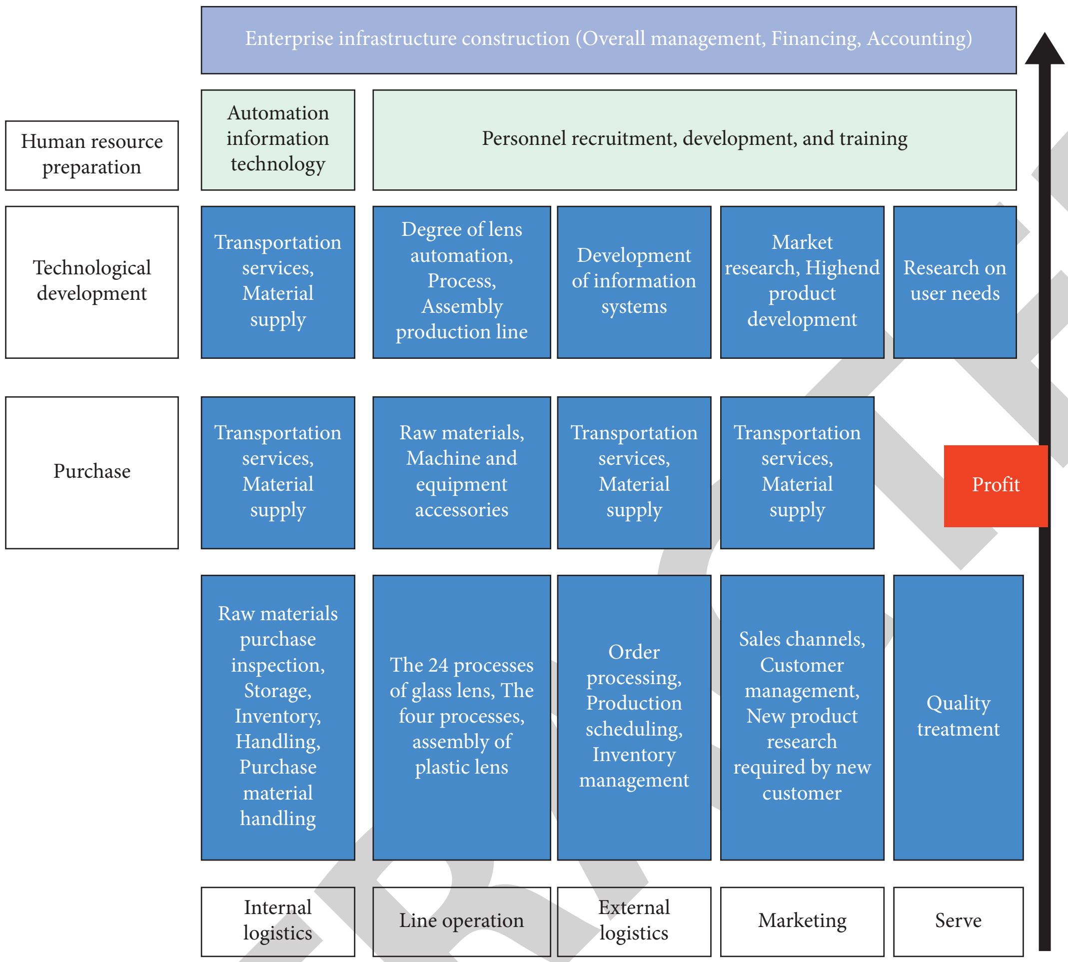


Figure 6: The value chain diagram of mobile phone companies.

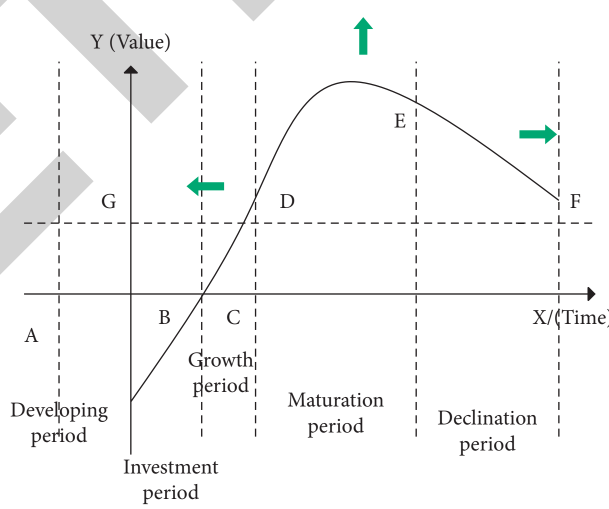


Figure 7: Life cycle characteristics of mobile phone products.

technology and production technology, and carry out secondary development on the basis of mastering core design technology and production technology. Finally, the direction of research and development needs to be adjusted. Domestic mobile phone companies should seize the opportunity of upgrading the 5G industry, accelerate the research and development, tracking, technical reserves, and other preparations for the third generation of mobile communications, master core technology as soon as possible, and completely change China's passive situation of

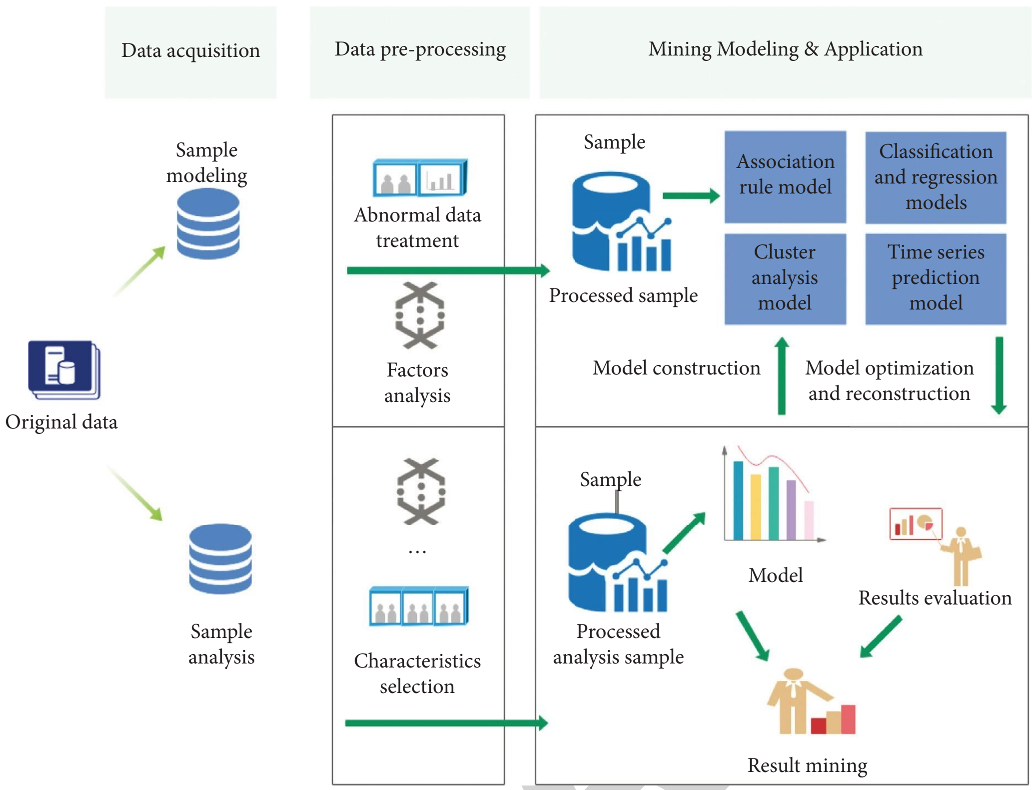


FIGURE 8: Flow chart of data mining modeling.

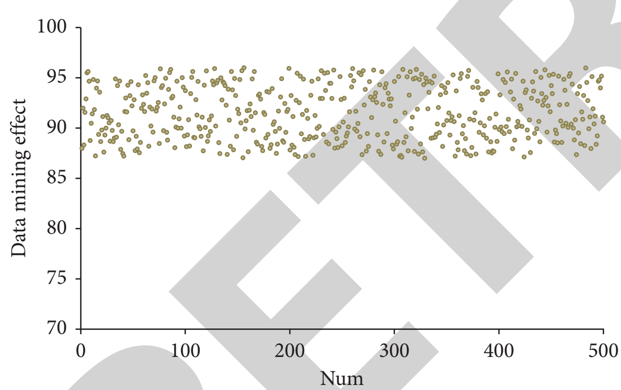


FIGURE 9: Cluster evaluation diagram of data mining effect.

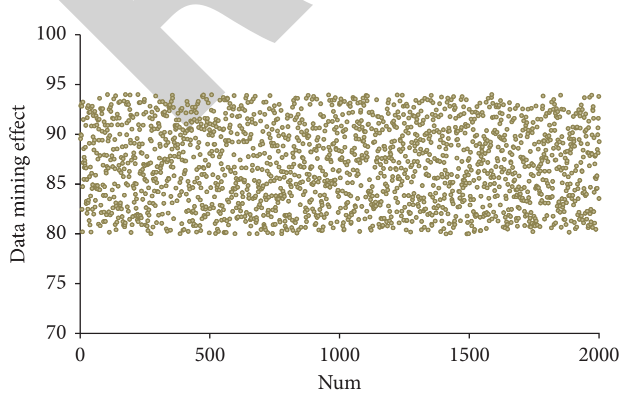


FIGURE 10: The clustering effect of the system in the analysis of the international competitiveness of the mobile phone industry.

being completely controlled by others in the global mobile phone industry chain.

5. Conclusion

As consumers' understanding of mobile phones continues to increase, they have become very rational in buying mobile phones and they have paid much attention to the value of products and services provided by enterprises. By consulting the related literature, this article combines the current status of China's mobile phone market research to identify the customer value evaluation indicators of mobile phone companies and then evaluates and scores various indicators that drive customer value through questionnaire research. Moreover, this paper assigns a certain weight to each indicator based on the data of the questionnaire survey, thereby calculating the customer perceived value of mobile phone companies in the Chinese market. In addition, this paper conducts a comparative analysis to obtain the comparative competitive advantages and disadvantages of various mobile phone companies and then proposes corresponding customer value creation strategies based on this. These have very important practical guiding significance for enhancing the competitiveness of domestic mobile phone companies. This article combines data mining algorithms to analyze the international competitiveness of China's mobile phone industry, increase the international market share of domestic mobile phones, and provide a reference for the subsequent development of China's mobile phone industry.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9870123, 1 page
https://doi.org/10.1155/2023/9870123

Retraction

Retracted: Optimization of Ideological and Political Education Management Strategies under *k*-Means Algorithm in Big Data Environment

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Wang, "Optimization of Ideological and Political Education Management Strategies under *k*-Means Algorithm in Big Data Environment," *Security and Communication Networks*, vol. 2022, Article ID 6120230, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6120230, 14 pages
https://doi.org/10.1155/2022/6120230

Research Article

Optimization of Ideological and Political Education Management Strategies under k-Means Algorithm in Big Data Environment

Hongyan Wang

School of Automotive Engineering, Henan Polytechnic Institute, Nanyang, Henan 473000, China

Correspondence should be addressed to Hongyan Wang; 2013012@hnpi.edu.cn

Received 9 February 2022; Revised 21 March 2022; Accepted 28 March 2022; Published 23 April 2022

Academic Editor: Chin-Ling Chen

The development of information technology has promoted the reform of ideological and political education in colleges and universities. A large amount of data has been accumulated in the education management database, and the information implied by these data can provide scientific guidance for the optimization of educational management strategies in colleges and universities. The relationship and characteristics of big data in ideological and political education in colleges and universities were expounded, and the feasibility of applying big data technology to ideological and political education was analyzed. The k-means algorithm was selected for cluster analysis, and its process and principles were expounded. As the traditional k-means clustering algorithm has low data processing efficiency and large deviation of the results, the algorithm was optimized by controlling the iterative method of the algorithm. Besides, the ideological and political education management under the optimized k-means algorithm was established. The work assessment quantitative scale in the management of ideological and political education was adopted as the data source, and the optimized k-means algorithm was used to carry out cluster analysis. The results show that management attitude was scored as 0.634, the management ability was 0.6092, the management effect was 0.6082, and the management method was 0.5792. It was indicated that all the scores were above the middle for greater than 0.5, suggesting that the overall management level was above the middle, which was relatively good. The optimized k-means-based ideological and political education management strategy model can analyze the current educational management status of colleges and universities more accurately. It can also provide scientific guidance for colleges and universities to conduct teaching management reasonably and scientifically according to the data analysis results. The optimized k-means algorithm was compared with the traditional algorithm, from which the optimized algorithm was obviously better than the traditional algorithm in terms of clustering effect and operation stability.

1. Introduction

With the continuous development of the electronic information industry, big data technology has also played a great role in all walks of life [1, 2]. It not only changes the way people live and learn, but also updates people's thinking constantly. In colleges and universities, the development of information technology has also promoted the reform of teaching methods and the optimization of teaching management strategies. Colleges and universities have gradually entered the era of large-scale data mining (DM) and data application [3]. How to use new media and new technologies to strengthen and innovate the ideological and political education is an important and realistic topic for colleges and universities in the new era [4, 5]. In the education and teaching management, the informatization of the teaching system and the education management system has led to the generation of massive amounts of data in various educational systems. There is a lot of information within these data. The application of DM technology in the education management system is of important theoretical significance and practical application value [6, 7]. Many scholars have conducted related in-depth works on the massive information contained in big data in a variety of educational systems.

Ji et al. (2020) [8] explored educational DM, with special attention to education big data mining algorithms. First, the relevant elements of educational DM were analyzed, and big data technology was introduced according to the needs of educational data application. Then, the commonly used education big data mining algorithms and their applications were also introduced, and finally, the development trend of education big data mining algorithms was discussed. Alsuwaiket et al. (2020) [9] expressed the extracted knowledge in a way ensuring accurate and reliable results. The student attendance data collected from the education system were cleaned up, to eliminate any randomness and noise. Then, various attributes were explored to highlight the most important attributes that the affected actual attendance of students. With the attributes selected, an equation was derived to measure the credibility of student attendance. The credibility of the newly developed measurement was also evaluated to check its consistency. Finally, the modules were classified in line with the strength of the attendance credibility value using the J48 DM classification technology. The result showed that the credibility value obtained by the derived equation gave an accurate, credible, and true index of the student attendance rate. The accurate classification of the modules was also performed regarding the credibility of the student attendance rate on these modules. Tan and Lin (2021) [10] proposed a new predictive model to detect the technical aspects of teaching and e-learning using DM in the virtual education system. The association rule mining and supervision technology were applied for factor detection in the virtual education system. The experimental results showed that the proposed prediction model satisfied the accuracy, precision, and recall factor in predicting student teaching and online learning behaviours in the virtual education system. Fang and Lu (2021) [11] analyzed learners' behaviour data in the learning process through the learner model, and three characteristics of learners' cognitive ability, knowledge level, and learning preference were extracted. The preference model was constructed by using the ontology, the semantic relationship among knowledge was better understood, and the interest of students in learning was discovered. Black et al. (2021) [12] analyzed the academic performance and behaviour of some engineering students and collected data from score tables and other related factors. The final model for two datasets was constructed under decision trees and naive Bayes algorithms, and the model could be used to predict the performance of students accurately.

From the current research, all kinds of education, whether psychological education, medical education, or learning model, all contain a lot of data mining and analysis. The management of ideological and political education in colleges and universities also contains a lot of data. Therefore, the ideological and political education of university A in China is taken as an example, the relationship and characteristics of big data in the ideological and political education of universities are first described, and the feasibility of the application of big data technology in the ideological and political education of universities is analyzed. Secondly, the clustering analysis is introduced in data mining, the process and principle of clustering analysis are described, and the k-means algorithm is selected in the partition method for clustering analysis. On this basis, the management strategy of ideological and political education based on the k-means algorithm is established. The Work Assessment Quantification Table in ideological and political education management in colleges and universities is taken as the data source, and the k-means algorithm is applied to cluster analysis.

2. Theoretical Basis

2.1. Big Data and Ideological and Political Education in Colleges and Universities

2.1.1. Big Data. As network technology develops rapidly and the network infrastructure has been improved continuously, data have become the most important method of information storage and information transmission in modern society. Especially with the rapid development of computer technology, the Internet of Things and the Internet have also been applied successively, which will undoubtedly generate countless data. Therefore, modern society has also entered the era of big data [13, 14]. Abundant information is hidden behind the data; the hidden information can be mined through DM and data analysis, thereby providing great guidance for work, research, and even social development. Firstly, the amount of big data is extremely huge, and the data are integrated. In modern society, everyone can make full use of data technology in life, and everyone is a producer of big data. The exchange of data follows new technologies. After the communication among people is increased for their work, research, and life, the use will also inevitably increase. With data development, the accuracy required by data is getting higher and higher. People can participate in the collection and mining of large amounts of data according to their own needs, and combine qualitative and quantitative analyses to explore the development of society and nature. Secondly, there are many types of data, which mainly include the structured data and unstructured data. People are moving into unstructured data, careers, working lives, research, and unstructured models, because most of the data are tied to human activity. Thirdly, the data processing block determines the processing speed of the data. The growth speed of the data is exponential. Therefore, it is necessary to process the data in a timely and effective manner. The data changes dynamically. If the value contained in the data is not processed and calculated timely, the data may lose its original functions, and big data itself becomes useless [15].

2.1.2. Characteristics of Ideological and Political Education in the Era of Big Data. Big data has become a keyword in modern society. For the ideological and political education in colleges and universities, there are a mass of data. It is necessary to make full use of data and share data to make ideological and political education more scientific [16, 17]. Therefore, ideological and political education in the big data age mainly has the following characteristics. (1) The main body of ideological and political education is surrounded by big data information. With the global diversification, different cultures and social systems of various countries can be displayed through big data. Different countries have different ideas, and the opposite sex have different cultural backgrounds and adopt specific solutions to different problems. In addition, due to the increasing dependence of

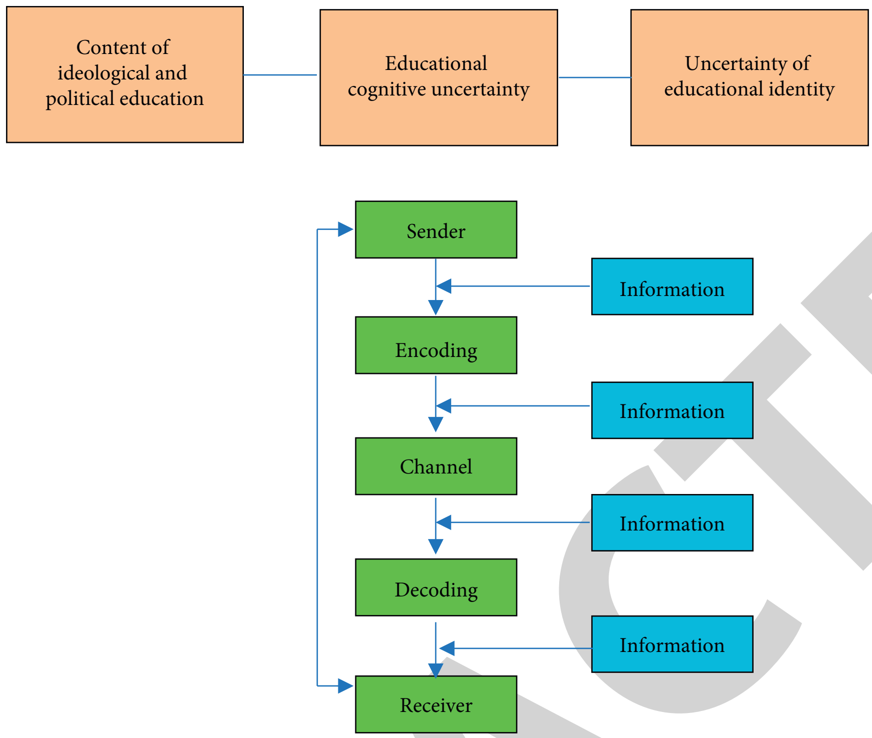


Figure 1: Information essence and information flow module.

personal life on data, the massive overloading information will have a deconstructive effect on the core ideas in China, impacting mainstream values and beliefs. (2) The object of ideological and political education is affected by big data information [18]. In ideological and political education, the most obvious word is indoctrination. The transmission of some ruling ideas will help to construct core values in society; but in the era of big data, the amount of logical information that can be contacted by the objects may be at the same level with that by educators. Educators are no longer the only indoctrination subjects [19]. Some communication platforms and mainstream media influence the minds of the educated. These data are not all positive. Thus, if there is no theoretical guidance, institutional constraints, and ethical guidance, it may cause confidential disclosure and chaos when using big data [20]. (3) The process of ideological and political education is disturbed by the dissemination of big data information [21, 22]. However, under the big data model, the huge amount of information and the characteristic of high dispersion can increase the uncertainty of information easily. There is too much valueless information increasing. In the ideological and political education, the information flow can also refer to the general information flow model, which is shown in Figure 1 [23].

2.1.3. The Relationship between Big Data and Ideological and Political Education. The relationship between big data and ideological and political education in colleges and universities is shown in Figure 2.

It could be observed from Figure 2 that in the ideological and political education in colleges and universities, the functions of big data mainly included information collection and screening, scientific prediction and judgment, and personalized education and guidance. Big data contains a lot of information. Some valuable information and negative information are usually mixed. Therefore, in college education, educators must collect data and screen data to filter out valuable information. The harmful or even undesirable information should be resisted. Educators could use big data technology to collect registration information and Internet activities of online users, and filter out the identities of university network users. Then, the core values and the character education of college students could be promoted and strengthened. Big data can be used to predict the network behaviour of college students and explore the law of ideological development of college students. In this way, problems can be found and effective ideological and political education can be carried out. In addition, in a diversified environment, the ideological and political status of different college students can be summarized through the conclusion of data visualization, and it becomes a quantitative and visual form. Thus, the conclusions could be drawn accurately, to conduct ideological and political education concretely for better guidance [24]. Big data technology was also applied to develop personalized hidden education for college students.

2.2. Clustering Technology in Big Data Mining. DM [25] refers to extracting knowledge that people are interested in from massive data. These information and knowledge are usually implicit and contain a lot of potential information. The mined information and knowledge can be expressed as

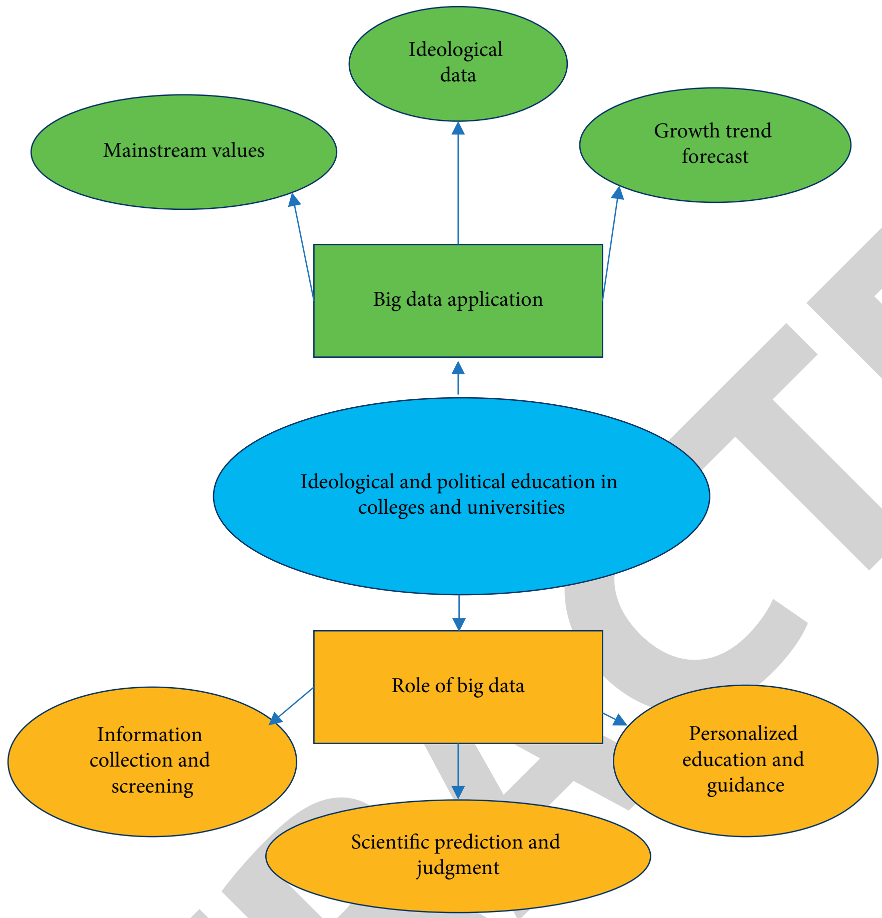


Figure 2: The relationship between big data and ideological and political education in colleges and universities.

concepts, rules, laws, and other visualized forms. DM is the process to seek for optimal decision support of the module in the plentiful information. Data obtained by DM can be classified into structured data and unstructured data regarding different objects [26, 27]. DM is a multidisciplinary technology. It is not just a simple data query, but is to dig out the hidden knowledge and information in the massive data from a low level to a high level. After further processing, the mined data can be directly provided to decision-makers to assist the decision-making process. Otherwise, experts can modify their existing body of knowledge. Or, it can be viewed as a knowledge storage structure that transforms new knowledge into an application system, such as expert systems and rule bases. Clustering technology is important in DM, and the application of clustering technology is quite extensive [28].

2.2.1. Cluster Analysis of the k-Means Algorithm. Cluster analysis is a process of dividing and grouping datasets of fictitious or physical data objects [29]. Its objective is to make similar data objects form a data set. During the process, a generated group of data objects is called a cluster, which is a collection of data objects. In a cluster, objects have a high similarity with each other. While in different clusters, there are large differences among objects. The application fields of cluster analysis are shown in Figure 3 [30].

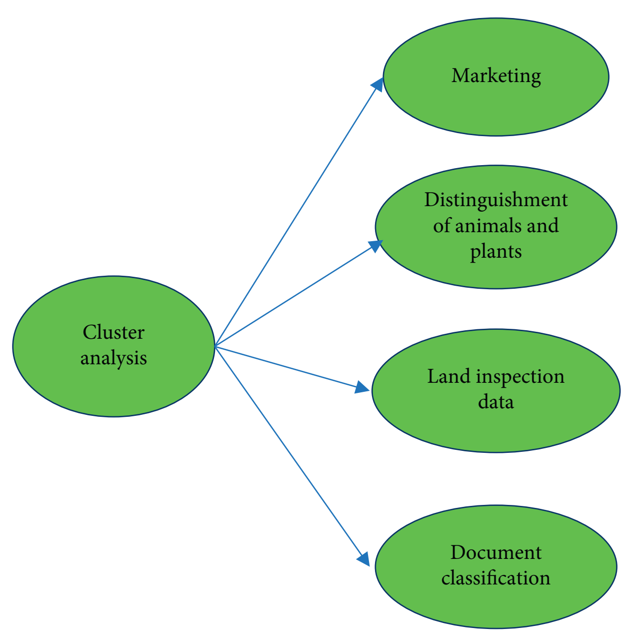


Figure 3: Application areas of cluster analysis.

It can be found from Figure 3 that cluster analysis has a wide range of applications. The in-depth analysis and mining can be conducted of data from different industries, to guide practice better. Now, the massive data sets are processed with the large and complex data warehouses, and DM puts forward higher requirements on the computing power of clustering algorithms.

2.2.2. Data Types of the k-Means Algorithm. In this section, more data types used in the clustering process are further discussed, thereby further analyzing the preprocessing of these data types [31, 32]. It was assumed that there were n data objects to be processed in the clustering issue. n objects represented different types, such as people, cars, materials, and so on. Generally, there are two types of data that appeared in cluster analysis.

2.2.3. Data Matrix. P variables were selected to describe n objects. For example, price, use, shelf life, and quality were used to describe the properties of the objects. Then, these data were measured with the interval scale to obtain an $n \times p$ matrix as shown in the following:

$$\begin{pmatrix} x_{11} & \mathrm{x}_{12} & \cdots & \mathrm{x}_{1\mathrm{p}} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{2\mathrm{p}} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}. \quad (1)$$

2.2.4. Dissimilarity Matrix. The dissimilarity matrix was used to place the similarity between two of n objects. Its specific expression can be described as an $n \times n$ matrix, which is shown in the following equation:

$$\begin{pmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \cdots & \cdots & 0 \end{pmatrix}. \quad (2)$$

In the above matrix, d(i,j)was the specific quantified form of the dissimilarity between two objects i and j. Generally, $\mathrm{i},j \geq 0$. When d(i,j) approached 0, it meant the similarity between the objects i and j was high. When d(i,j) approached the maximum value, it indicated that the dissimilarity between i and j was high. Usually, there were different measurement methods for the dissimilarity d(i,j). At present, a lot of clustering analyses are based on the dissimilarity matrix, but in some cases, the formation of data is described by the data matrix, so the data matrix should be converted into the dissimilarity matrix before clustering algorithm analysis. The next step is discussing the measurement method of dissimilarity d(i,j), usually using the following variable types.

a. Interval scale measurement. The degree of dissimilarity (or similarity) between objects was described by interval scale variables, which was measured by the distance between the objects. The most classic method of calculating the distance is Euclidean distance, which is defined as follows:

$$d(i,j) = \sqrt{\left|x_{j1} - x_{j1}\right|^2 + \left|x_{i2} - x_{j2}\right|^2 + \cdots + \left|x_{in} - x_{jn}\right|^2}. \quad (3)$$

Here, $i = (x_\lambda, x_{i2}, \cdots, x_{in})$ and $j = (x_\lambda, x_{j2}, \cdots x_{\overline{\mu m}})$ were n-dimensional data objects. Another distance measurement method is Manhattan (or city block) distance, which is defined as

$$d(i,j) = \left|x_{i1} - x_{j1}\right| + \left|x_{j2} - x_{j2}\right| + \cdots + \left|x_{im} - x_{jN}\right|. \quad (4)$$

Both Equations (3) and (4) needed to meet the following conditions. When $\mathrm{d}(i,j) \geq 0$, both the distances were nonnegative values. When $\mathrm{d}(i,j) = 0$, the distances were 0. When $\mathrm{d}(i,j) = d(i,j)$, the function between the two distances had a symmetry $\mathrm{d}(i,j) \leq d(i,\mathrm{h}) + d(\mathrm{h},j)$ suggested that the direct distance from object i to object j was less than or equal to the other distance to object j through object k.

b Binary variable. Each variable had two states, which were represented by 0 and 1. 0 means it is empty, and 1 means it is present. It was assumed that the weight of any binary variable was the same, and then, a 2×2 conditional dependency table could be obtained, which is shown in Table 1.

Table 1: Condition dependency table.

i/j	1	0	Sums
1	q	r	$q+r$
0	s	t	$s+t$
Sums	$q+s$	$r+t$	P

In Table 1, q represents the number of binary variables where the object i was 1, and r represents the number of binary variables where the object i was 1 when the object j was 0. s was that where the data object i was 0 and the data object j was 1, and t was that where both the object i and object j took 0. P was the total number of binary variables, and $p = q + r + s + t$.

A simple matching correlation coefficient was used to describe the dissimilarity between object i and object j, which could be defined as

$$d(i,j) = \frac{(r+s)}{(q+r+s+t)}. \quad (5)$$

If the meanings included in the two states of 0 and 1 were not equally important, the binary variable was called an asymmetric binary variable. In such a case, the Jaccard coefficient was the rating method for measuring the dissimilarity of the asymmetric binary variable. It is expressed as

$$d(i,j) = \frac{(r+s)}{(q+r+s)}. \quad (6)$$

Categorical variables are further extensions of binary variables. For instance, the categorical variable blood type has four possible state values, namely, O-type, A-type, B-type, and AB-type. The categorical variables i and j could be calculated with dissimilarity between them, which is expressed as

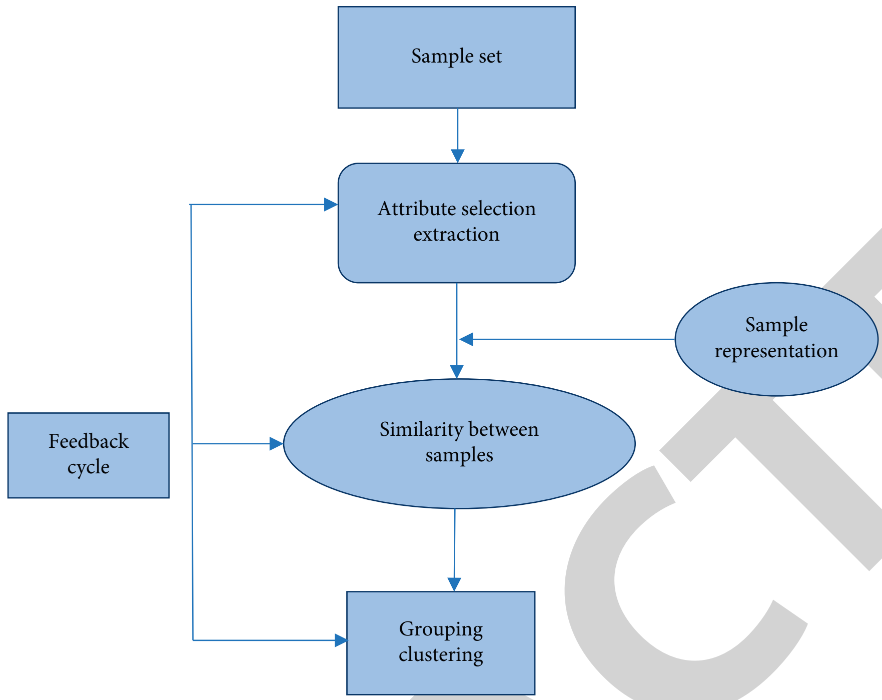


Figure 4: Process of cluster analysis.

$$d(i,j)=\frac{p-m}{p}. \tag{7}$$

Here, m is the number of matches and p is the number of all variables.

2.3. Processes of Cluster Analysis and k-Means Algorithm

2.3.1. Cluster Analysis Process. Figure 4 shows the flow of the cluster analysis.

It was found from Figure 4 that the input end of cluster analysis was a sample set composed of data that needed to be processed. The sample set data were preprocessed and classified, and then, the data attributes were extracted. The clusters are grouped according to the similarity of the samples. Cluster analysis was usually a cyclic process, and finally, a sample cluster was obtained. In the clustering algorithm, there are many partition methods. If the number of data objects to be processed was set as n, the partition method firstly adopted is the objective function minimization. With repeated iterations, the n objects are divided into k blocks. Each block represents a cluster, and k is much smaller than n. There are many partition methods, among which the k-means algorithm was the most commonly used.

2.3.2. k-Means Algorithm. The number of data objects to be processed was n. The division method was the strategy of minimizing the objective function. Through repeated iteration, the n objects were divided into k blocks, each block represented a cluster, and $k << n$ was satisfied. In other words, the k divisions must meet the following two conditions: (1) each cluster had at least one object. (2) Each object corresponded to a cluster. k-Means is a commonly used division method, so the algorithm was focused on [33].

The k-means algorithm was to calculate the mean value of the objects in each cluster, which was just the centroid in the cluster. The processing steps of k-means were as follows. k objects were arbitrarily selected as the preset k centroid, and then, the centroid of each new cluster was calculated. The above steps were repeated until the criterion function converged to minimize the objective function. The commonly used criterion function was the squared error criterion function, which is shown in the following equation:

$$E=\sum_{i=1}^{t}\sum_{p\in c_j}\left|p-m_j\right|^2. \tag{8}$$

Here, E represents the sum of the squared errors of all objects in the space; p denotes all points in the dataset, used to represent the data objects. m_j represents the centroid of the cluster c_j, where both p and m_i are multidimensional. The k-means algorithm flow is shown in Figure 5; it aimed at making the E value smaller and smaller. When the final clusters were compact but distinguishable with each other, it indicated that the algorithm performed well. Especially when the number of objects in the dataset was large, the efficiency of the k-means algorithm was better, because its complexity could be calculated as o (nkt), in which n represents the number of objects in the dataset and t is the number of iterations. In general, k was much smaller than n, and t was also much smaller than n. The clustering stopped when a local optimum was obtained. The following was an example to illustrate the k-means algorithm.

5 points $\{x_1, x_2, x_3, x_4, x_5\}$were described by two-dimensional coordinates, which represent the two-dimensional samples of the cluster analysis $x_1=(0,4)$, $x_2=(0,0)$, $x_3=(3,0)$, $x_4=(4,0)$, and $x_5=(5,3)$. The number of initial clusters k

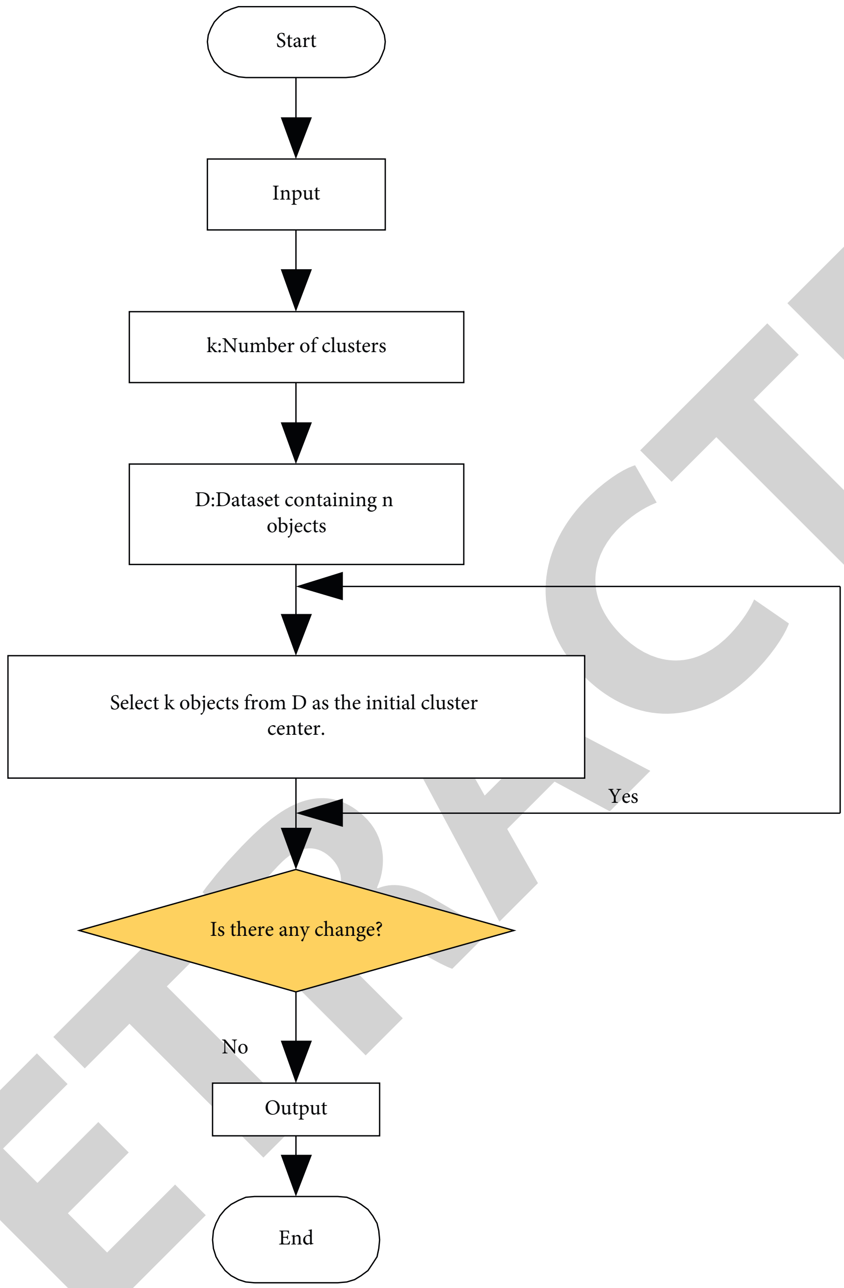


Figure 5: Flowchart of k-means algorithm.

was 2. Then, the k-means algorithm was carried out to analyze the following.

Step 1. k objects were arbitrarily selected from the given data samples and taken as the initial cluster centers. The centers M_1 and M_2 of these two clusters are expressed in

$$\begin{aligned} M_1 &= x_1 \\ &= (0, 4). \end{aligned} \tag{9}$$

$$\begin{aligned} M_2 &= x_2 \\ &= (0, 0). \end{aligned} \tag{10}$$

Step 2. For the rest of the data objects, the distances from the cluster centers M_1 and M_2 to them were computed using Euclidean distances. Each object was also reassigned to the closest cluster.

For x_3, the following is obtained:

$$\begin{aligned} d(M_1, x_3) &= \left((0-3)^2 + (4-0)^2\right)^{\frac{1}{2}} = 5 \\ d(M_2, x_3) &= \left((0-3)^2 + (0-0)^2\right)^{\frac{1}{2}} = 3 = \Longrightarrow x_3 \in c_2. \end{aligned} \tag{11}$$

For x_4, the distances are computed in the following:

$$d\left(M_1, x_4\right) = \left((0-4)^2 + (4-0)^2\right)^{\frac{1}{2}} = 5.7$$
$$d\left(M_2, x_4\right) = \left((0-4)^2 + (0-0)^2\right)^{\frac{1}{2}} = 4 = \Rightarrow x_4 \in c_2. \tag{12}$$

For x_5, the distances are calculated in the following equation:

$$d\left(M_1, x_5\right) = \left((0-5)^2 + (4-3)^2\right)^{\frac{1}{2}} = 5.1$$
$$d\left(M_2, x_5\right) = \left((0-5)^2 + (0-3)^2\right)^{\frac{1}{2}} = 5.8 = \Rightarrow x_5 \in c_2. \tag{13}$$

They were updated, and the new clusters $c_1 = \{x_1, x_5\}$and$c_2 = \{x_2, x_3, x_4\}$ were obtained.

The squared error criterion was calculated, and the mean value of the objects in each cluster is worked out by the following equation:

$$\begin{aligned} E_1 &= \left[(0-0)^2 + (4-4)^2\right] + \left[(0-5)^2 + (4-3)^2\right] = 26 \\ E_2 &= \left[(0-0)^2 + (0-0)^2\right] + \left[(0-3)^2 + (0-0)^2\right] \\ &\quad + \left[(0-4)^2 + (0-0)^2\right] = 25. \end{aligned} \tag{14}$$

The overall squared error was expressed as $E = E_1 + E_2 = 51$.

Step 3. The new centroid was calculated.

$$M_1 = \left(\frac{(0+5)}{2}, \frac{(4+3)}{2}\right) = (2.5, 3.5)$$
$$M_2 = \left(\frac{(0+3+4)}{3}, \frac{(0+0+0)}{3}\right) = (2.33, 0). \tag{15}$$

Steps 1 and 2 were carried out repeatedly. The distances between each point and the new centroid M_1 and M_2 in this step are calculated in the following equations.

For x_1, the equation for the distances is shown as follows:

$$d\left(M_1, x_1\right) = \left((2.5-0)^2 + (3.5-4)^2\right)^{\frac{1}{2}} = 2.55$$
$$d\left(M_2, x_1\right) = \left((2.33-0)^2 + (0-4)^2\right)^{\frac{1}{2}} = 4.63 = \Rightarrow x_1 \in c_1. \tag{16}$$

For x_2, the distances are computed as shown in the following:

$$d\left(M_1, x_2\right) = \left((2.5-0)^2 + (3.5-0)^2\right)^{\frac{1}{2}} = 4.3$$
$$d\left(M_2, x_2\right) = \left((2.33-0)^2 + (0-0)^2\right)^{\frac{1}{2}} = 2.33 = \Rightarrow x_2 \in c_2. \tag{17}$$

For x_3, the distances are obtained in the following equation:

$$d\left(M_1, x_3\right) = \left((2.5-3)^2 + (3.5-0)^2\right)^{\frac{1}{2}} = 3.8$$
$$d\left(M_2, x_3\right) = \left((2.33-3)^2 + (0-0)^2\right)^{\frac{1}{2}} = 0.67 = \Rightarrow x_3 \in c_2. \tag{18}$$

For x_4, the following equation calculates the distances:

$$d\left(M_1, x_4\right) = \left((2.5-4)^2 + (3.5-0)^2\right)^{\frac{1}{2}} = 3.8$$
$$d\left(M_2, x_4\right) = \left((2.33-4)^2 + (0-0)^2\right)^{\frac{1}{2}} = 1.67 = \Rightarrow x_4 \in c_2. \tag{19}$$

For x_5, the following equation computes the distances:

$$d\left(M_1, x_5\right) = \left((2.5-5)^2 + (3.5-3)^2\right)^{\frac{1}{2}} = 2.55$$
$$d\left(M_2, x_5\right) = \left((2.33-5)^2 + (0-3)^2\right)^{\frac{1}{2}} = 4.02 \Rightarrow x_5 \in c_1. \tag{20}$$

The new clusters $c_1 = \{x_1, x_5\}$and $c_2 = \{x_2, x_3, x_4\}$ of this step were obtained as these above were updated. The centroids were $M_1(2.5, 3.5)$ and $M_2(2.33, 0)$.

The corresponding variance and squared error are expressed in the following equation:

$$\begin{aligned} E_1 &= \left[(0-2.5)^2 + (4-3.5)^2\right] + \left[(2.5-5)^2 + (2-3)^2\right] \\ &= 13.75 \\ E_2 &= \left[(0-2.33)^2 + (0-0)^2\right] + \left[(3-2.33)^2 + (0-0)^2\right]. \\ &\quad + \left[(4-2.33)^2 + (0-0)^2\right] = 8.67 \end{aligned} \tag{21}$$

The overall squared error in this step was described as $E = E_1 + E_2 = 22.42$.

It could be clearly shown from the above steps that after one iteration, the overall error decreased a lot from 51 to 22.42. If repeated iterations were continued, the same conclusion would be given. The reason was that all the samples would be divided into the same cluster; as they would not be redivided, the algorithm ended and it will not change any longer.

2.4. Optimization and Improvement of k-Means Algorithm. Since the selection of the cluster center directly affected the results of the clustering, and average error of many clusters was relatively large, the local optimum in the clustering process occurred. Therefore, in the process of each iteration, the weight of each data object was regulated, the cluster center was recalculated, and the threshold was adjusted. In addition, the method of weighting was introduced to measure the size of the weight assigned to the cluster by controlling the error value. Through the weighted penalty method, the data objects farther from the cluster center in the cluster with larger average error would be reclassified to the adjacent cluster with smaller average error. In this iterative process, the boundaries of the clusters were continuously determined by weighted thresholds, until the data objects were classified into the correct clusters. The penalty factor and the memory factor were introduced, and the penalty control of the cluster was realized through the changes of these two factors. The memory factor was initialized by the influence of the weight of the last iteration on the weight of the cluster in this iteration. Then, the penalty factor was introduced to change the cluster with no weight change in the last iteration, so as to improve the stability of the algorithm. Finally, the clustering results could converge to the global optimum successfully. The specific optimization method was described as below.

The dataset to be clustered was defined. At the end of the t-1-th iteration, the k cluster centers are expressed as

$$C = \left\{C_m^{t-1}\right\}_m^{t-1}. \tag{22}$$

After the t-1-th iteration, the k clusters are represented as follows:

$$V = \left\{V_m^{t-1}\right\}_{m-1}^{\mathrm{k}}. \tag{23}$$

The error value of the cluster is denoted as follows:

$$f = \left\{f_m^{t-1}\right\}_{m-1}^{\mathrm{k}}. \tag{24}$$

Then, the penalty factor P_{k}^t of cluster k at the t-th iteration was defined by the iteration result of the t-1-th iteration. It was expressed in the following equation:

$$P_{\mathrm{k}}^t = \frac{f}{1 + \mathrm{esp}\,(V - C)}. \tag{25}$$

The memory factor was introduced into the weight update process, and the cluster weight calculation result of the last iteration was substituted. As the memory factor was set as α, the weight of cluster k at the t-th iteration was defined in the following equation after being improved:

$$W_{\mathrm{k}}^t = \alpha P_{\mathrm{k}}^t + (1 - \alpha). \tag{26}$$

3. Path Exploration for Big Data to Promote Ideological and Political Education

3.1. Opportunities and Challenges of Ideological and Political Education in Big Data. The statistics has been made on the characteristics of big data. The data of ideological and political education came from the 2021 quantification scale for the work assessment of counselors in a university in China. The general characteristics of big data are shown in Figure 6.

Figure 6 shows that the amount of big data was huge, the data were integrated, and there are many types of data. The data mainly included unstructured data, which was in a very important position, regardless of the total amount and the reliability of the analysis source.

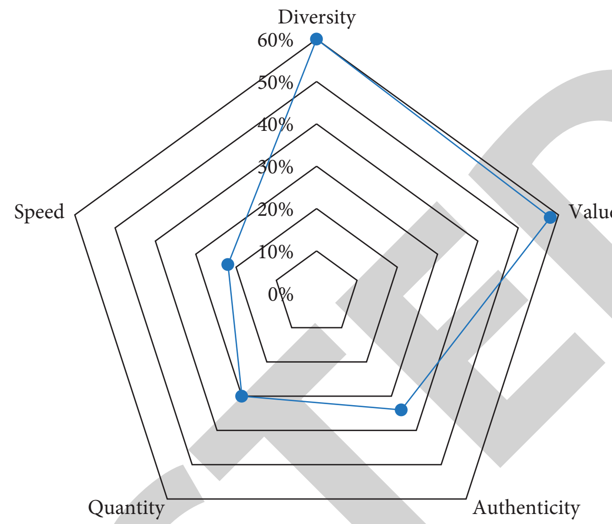


Figure 6: Feature attention of big data.

3.1.1. Opportunities of Big Data in Ideological and Political Education. Big data has opened the vision of the subjects of ideological and political education in colleges and universities, changes the traditional education mode, and makes the boundary between the objects and the subjects of education no longer strictly clear. It also impacts the thinking of the educational subjects and breaks the boundary of space. Through high-tech sharing of data, the convenience of technology is widely used in ideological and political education, which truly realized the integration of resources, differentiated analysis, and specific practical teaching. Secondly, big data can improve the quality of ideological and political education objects in colleges and universities. Data expand the knowledge of college students, and students can satisfy their curiosity through data. Students conduct DM, data collection, data analysis, and induction, forming their knowledge structure. Thirdly, big data has updated the dissemination way of ideological and political education in colleges and universities. In traditional information transmission, letters were the main form of communication. However, due to time and space limitations, the information was not transmitted timely. The resource sharing of big data is completely free of time and space constraints. Students can receive school news, regulations, policies, notices, and information freely, and can also strengthen their interaction with the colleges.

3.1.2. Challenges of Big Data in Ideological and Political Education. Big data has brought about tremendous changes in ideological and political education, but meanwhile, there are some challenges inevitably. The first is information asymmetry, which indicates differences in the information and structure of the two parties receiving the information. Secondly, big data threatens personal privacy and freedom.

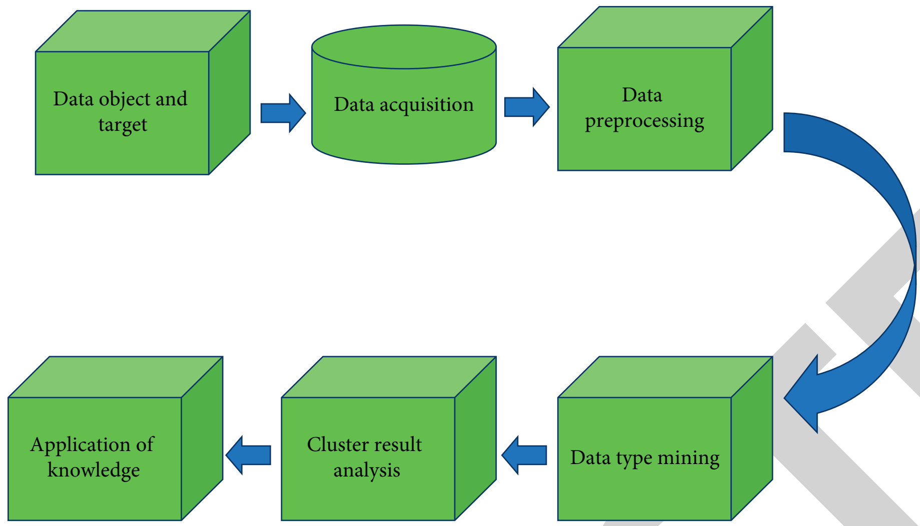


FIGURE 7: Clustering DM process.

Some shopping software would record and count the shopping habits of viewers, and some social software would keep a large amount of personal private information in the background. Thirdly, the digital divide has caused the information gap to widen. At last, culture is left behind science and technology.

3.2. Construction of Ideological and Political Education Management Strategy under Optimized k-Means Algorithm. The application of clustering technology in ideological and political education was mainly explored. The counselor work assessment quantitative scale was adopted to conduct DM and analysis. Various data were preprocessed, and finally, cluster analysis was performed through the k-means algorithm.

3.2.1. Ideological and Political Education Management. The continuous deepening of the ideological and political education, as well as the expanding enrollment and other policies of major colleges and universities, has a greater impact on the smooth implementation of college teaching plans. Therefore, the issue between management and learning has always been a difficulty. Therefore, colleges and universities established a team of student counselors. It is very urgent for the current college management to construct a high-efficiency and well-controlled decision supporting system. The system consists of a plenty of data in the education management, but these data only stay at the low-level stage. That's to say, it is only a simple query and simple classification of the data, and the information contained in the data is not mined. Thus, the counselor work assessment quantization scale was taken to dig deeper into the information.

3.2.2. Solution. As the counselor work assessment scale was adopted, cluster analysis was applied to perform DM. The answers to some questions were obtained, such as "How is the management level of college counselors?", "What loopholes exist in management?", and "Which aspects of management are not in place?". The DM process is shown in Figure 7.

From Figure 7, it was suggested that DM for the counselor work assessment scale was divided into 6 steps. First, the data object and its target needed to be clarified. Mining data blindly usually failed to get the desired results instead of some invalid and useless information. Data collection was also the most onerous part. Some data could be obtained directly, and the other data needed to be obtained through investigation. In data preprocessing, because not all collected data could be used, there was a lot of redundant information in the data, and the data needed to be processed into the types available for the model. Data clustering mining was to classify the data, as the data were divided into different groups according to the data models. After clustering result analysis, and finally, the application of knowledge was performed.

3.2.3. Ideological and Political Management Plan under k-Means. According to Figure 7, a series of tasks such as data collection were carried out:

(1) Determination of the objects and goals of DM. The data in the counselor work assessment quantitative scale of university A in 2021 were collected. The scale included a series of questions such as "How is the management level of college counselors?", "What are the loopholes in management?", "What in the management are not in place?".
(2) The counselor work assessment scales in 2021 were collected from the Academic Affairs Office of University A.
(3) For data preprocessing, the collected data needed to be converted. With the four attributes of management attitude, management ability, management method, and management effect, the data in the work

assessment scales were reorganized and merged. Then, the four attributes were quantified, and the five grades of the evaluation rating are sorted in line with excellent, good, qualified, poor, and awful grades. The grades were projected to the interval [0, 1] using the mapping method. The quantized values of the five grades were set to be 1, 0.75, 0.5, 0.25, and 0, respectively. The values of the four attributes were measured by the arithmetic mean of the items included in them, which are described as follows.

Management attitude = (insistence on standards + harmonious relationship with students + decent talks + objectivity)/4.

Management ability = (accurate grasp of the situation of poor students + accurate grasp of the situation of special student groups + strict education and investigation of students who violate disciplines + competent work + earnest organization for students to do good work)/5.

Management method = (class and dormitory visits more than 3 times a week + insistence on checking the hygiene of the student dormitory once a week + prepared talks with the students during the school year + active participation in and check of the morning running of the students + student scholarships in place)/5.

Management effect = (active participation in or hosting of meetings + understanding of the situation after class + comments and suggestions)/3.

3.2.4. Algorithm Implementation. Python was used for further implement of the k-means algorithm. It was supposed that there was a dataset D = $\{X_1, \cdots, X_n\}$, which aimed at finding k clusters $\{C_1, \cdots C_A\}$. The algorithm steps are shown in Figure 8.

The algorithm flow chart is shown in Figure 9.

4. Result Analysis

The commonly used five clustering algorithms—k-means algorithm, clustering using representative (CURE) algorithm, balanced iterative reducing and clustering using hierarchies (BIRCH) algorithm, density-based spatial clustering of applications with noise (DBSCAN) algorithm, and statistical information grid (STNG) algorithm—are compared. The comparison focuses on six aspects, namely, the scalability of the algorithm, the detection of arbitrary shape clustering, the ability to process noisy data, the sensitivity to the order of input objects, the multidimension, and the efficiency of the algorithm. The comparison results are shown in Table 2.

Since clustering technology is adopted in many application fields of databases, and different application layers have different requirements for clustering algorithms. Table 2 can be used as a reference for selecting clustering algorithms.

In the clustering technology of small- and medium-scale databases, the partition method can get local optimal solution. The partition method has more advantages than other clustering algorithms in terms of easy understanding,

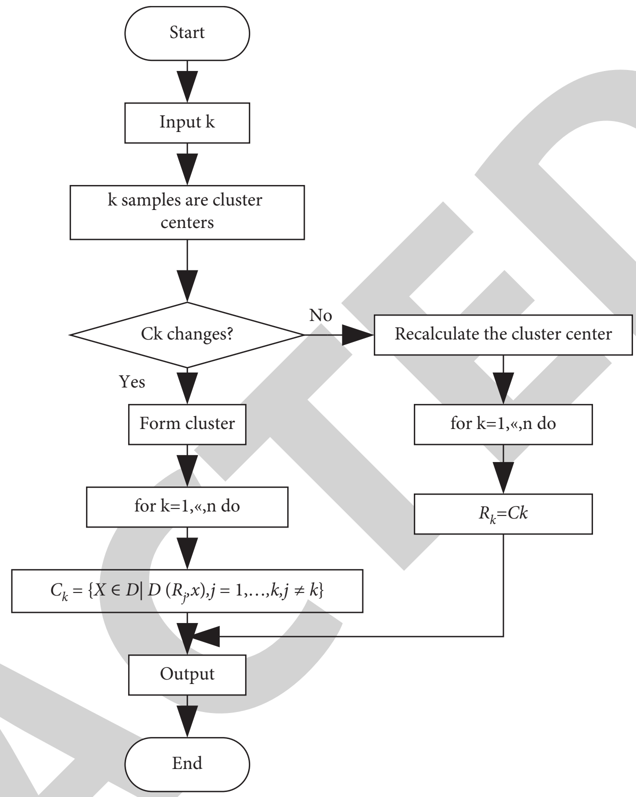


FIGURE 8: Algorithm steps.

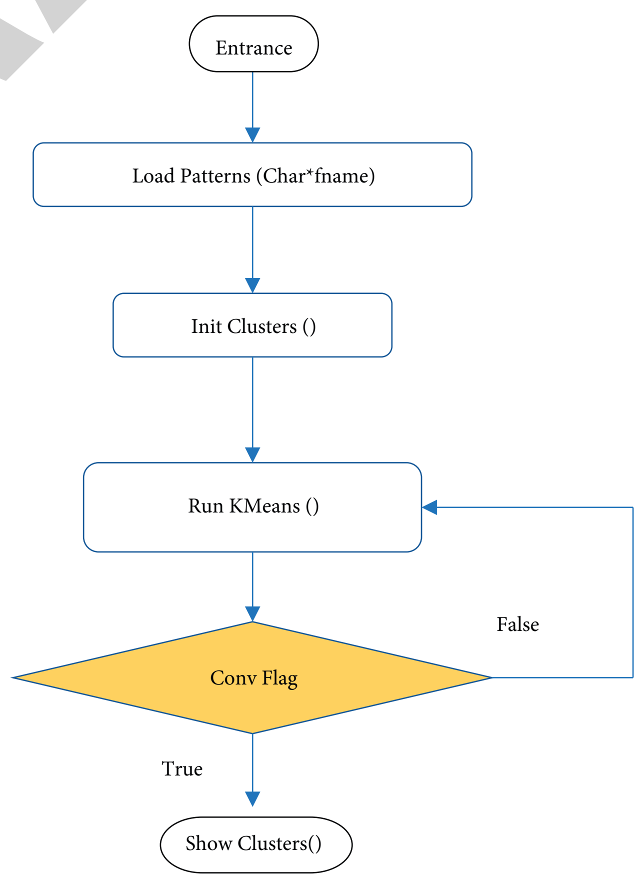


FIGURE 9: Flow chart of the algorithm.

Table 2: Comparison results of clustering algorithms.

Algorithm	Scalability	Detection of arbitrary shape clustering	Ability to process noisy data	Sensitivity to the order of input objects	Multidimension
k-Means algorithm	Great	Convex or globose	Sensitive	Sensitive	Medium
CURE	Poor	Arbitrary shape	Insensitive	Sensitive	Good
BIRCH	Good	Convex or globose	Medium	Medium	Good
DBSCAN	Very good	Arbitrary shape	Insensitive	Sensitive	Medium
STNG	Great	Arbitrary shape	Insensitive	Insensitive	Good

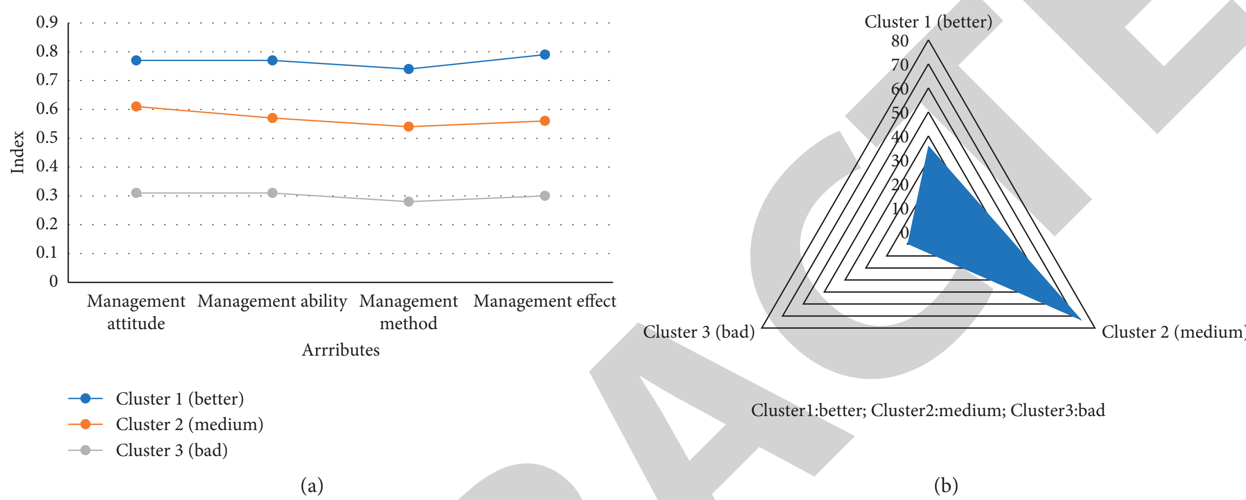


Figure 10: (a) Clustering results and (b) sample size.

easy training, easy implementation, and wide use. After the comparison of algorithms, the k-means algorithm is finally selected to achieve the classification of sample data.

The optimized k-means algorithm was applied to classify the sample data, and the three standard samples representing good, medium, and poor ones were compared, which is shown in Figure 10.

From Figure 10, the final proportional distribution range of the included data samples for each cluster was calculated. The proportional distribution range of cluster 1 (good) was 30%, that of cluster 2 (medium) was 62%, and that of cluster 3 (poor) was 8%. Given the results of DM again, the centroids of the three evaluation grades were carefully considered and were compared with the defined standard samples. Most of them were improved, except that the attribute value of the first cluster management method was 0.74, which was lower than the 0.75 of the defined standard sample. The overall score of other clusters was all increased. To get the overall score of each single attribute, weighting was performed due to the unequal sample sizes. The weighting coefficient was the proportion of the samples of the three grades assessed by clustering in the total samples, which were 30%, 62%, and 8%, respectively. The results are shown in Figure 11.

It was found from Figure 11 that the overall scores were ranked from high to low as management attitude (0.634), management ability (0.6092), management effect (0.6082), and management method (0.5792). It could be concluded

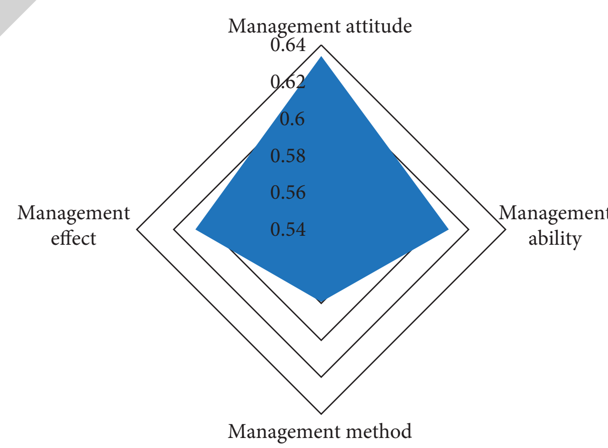


Figure 11: The overall scores of the four attributes.

that all the scores were in the upper-middle range as they were greater than 0.5, indicating that the overall management was in an upper-middle level. Thus, the management was relatively good. In the management of students, it was necessary to further strengthen the improvement of management methods, fully understand the actual ideological status of students, understand the life and learning of students actively, and improve the scholarship granting system.

The performance of the optimized k-means algorithm and the traditional algorithm was compared and analyzed. For the verification of the effectiveness and feasibility of the

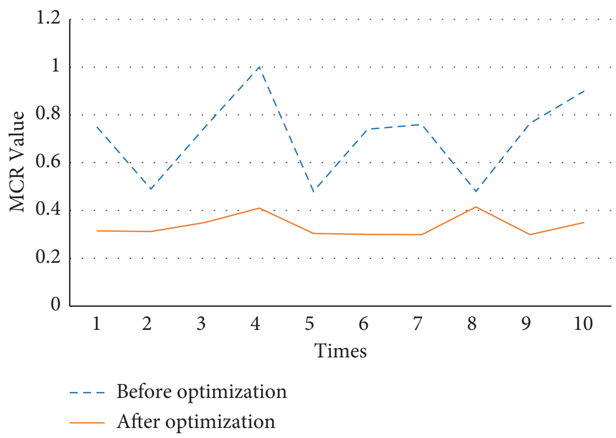


Figure 12: Comparison of MCR index values before and after k-means optimization.

improved k-means algorithm, Gaussian function was used to construct random data, and 600 two-dimensional data points were chosen as artificial datasets. The scale parameters were set as $\mu = 0.33$, $\mu = 0.87$, and $\mu = 1.6$. On this basis, the MCR (M indicates for the dataset, C the cluster, and R the Euclidean space) index was adopted to determine the effect of the artificial data clustering experiment. The MCR index was the ratio of the sum of the improved Euclidean distances to the amount of clustered data. The smaller the value of the MCR index, the higher the effectiveness of the k-means algorithm used. The simulation clustering was performed for 10 times continuously, and the obtained MCR index values are shown in Figure 12.

From the comparison in Figure 12, it was suggested that for the rough k-means clustering of different multidimensional cluster centers, the improved algorithm had a higher accuracy and a smaller average Euclidean distance than the common rough k-means algorithm. It was easy to classify the data into correct clusters, and the improved algorithm promotes the stability and feasibility of the algorithm to a certain extent.

5. Conclusion

College education has been into the era of big data. In the management of college education and teaching, there are massive data in various educational systems because of the informatization of the teaching system and the informatization of the educational management system. The traditional k-means clustering algorithm had a low efficiency in data processing with large deviation of the results; thus, the algorithm was optimized. The optimization model for the management strategy of ideological and political education in colleges and universities was also constructed under the optimized k-means algorithm in the context of big data. The work assessment quantitative scale in the management of ideological and political education in colleges and universities was taken as the data source, and the optimized k-means algorithm was used for clustering analysis. The clustering of the work assessment quantitative scale for the counselors was then realized. The students' comprehensively quantitative score reflected the counselors' work effectiveness directly and was to verify the clustering conclusion of the counselors' work assessment quantification model. It provided scientific guidance for the development of educational work according to the cluster analysis results. Compared with the traditional algorithm, the optimized algorithm was obviously better in terms of clustering effect and operation stability.

The traditional analysis method is based on the calculation of absolute scores, which has some defects in the objectivity and accuracy of evaluation results. It is unfair to evaluate counselors according to the traditional analysis method, and it cannot evaluate the management effect of counselors effectively and properly. Hence, cluster analysis is introduced into the application of ideological and political education management in colleges and universities. The application of data mining technology to automatic data analysis and mining of many evaluation data to find useful information can effectively overcome the defects and deficiencies of traditional analysis methods.

Data Availability

The labeled dataset used to support the findings of this study is available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] S. Tu and M. Zhang, "Research on planning and design of settlement from cities to rural areas based on big data technology," *Soft Computing*, pp. 1–12, 2021.

[2] K. Rybicka, "Usage of big data technology in controlling," *Research in World Economy*, vol. 10, no. 4, 92 pages, 2019.

[3] G. Du, Z. Liu, and H. Lu, "Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment," *Journal of Computational and Applied Mathematics*, vol. 386, no. 12, Article ID 113260, 2021.

[4] C. Hetrick, C. M. Wilson, E. Reece, and M. O. Hanna, "Organizing for urban education in the new public square: using social media to advance critical literacy and activism," *The Urban Review*, vol. 52, no. 1, pp. 26–46, 2020.

[5] A. Dave, "Review of the textbook and the lecture: education in the age of new media by norm friesen," *The Liminal: Interdisciplinary Journal of Technology in Education*, vol. 1, no. 1, p. 10, 2019.

[6] F. F. Bbosa, J. Nabukenya, P. Nabende, and R. Wesonga, "On the goodness of fit of parametric and non-parametric data mining techniques: the case of malaria incidence thresholds in Uganda," *Health Technology*, vol. 11, no. 4, pp. 929–940, 2021.

[7] H. Asami, M. Golabi, and M. Albaji, "Simulation of the biochemical and chemical oxygen demand and total suspended solids in wastewater treatment plants: data-mining approach," *Journal of Cleaner Production*, vol. 296, no. 2-4, Article ID 126533, 2021.

[8] L. Ji, X. Zhang, and L. Zhang, "Research on the algorithm of education data mining based on big data," in *Proceedings of the 2020 IEEE 2nd International Conference on Computer*

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1496170, 10 pages
https://doi.org/10.1155/2022/1496170

Research Article

Statistics and Analysis of Targeted Poverty Alleviation Information Integrated with Big Data Mining Algorithm

Meizhen Gao ,[1] Li Li,[1] and Yetong Gao[2]

[1]*Jiaozuo Normal College, Jiaozuo 454000, Henan, China*
[2]*Xi'an University of Technology, Xi'an 710048, Shanxi, China*

Correspondence should be addressed to Meizhen Gao; gmz1970@jzsz.edu.cn

Received 18 February 2022; Revised 8 March 2022; Accepted 18 March 2022; Published 23 April 2022

Academic Editor: Chin-Ling Chen

To gain a more comprehensive and systematic understanding of the impact of government assistance to poor households on poverty reduction targets, a targeted poverty alleviation information statistics and analysis integrated with big data mining algorithm is proposed. Combined with the big data knowledge of the new era, according to the machine learning (ML) pipeline module in spark, a big data computing framework, combined with known data mining algorithms, massive sample data are used to replace random stratified sampling data for modeling and analysis, and random forest model, logistic model, and newly proposed waterfall model are constructed for poor households. Finally, through the comparative evaluation of several poor household identification models, the results show that when 100 real data test the accuracy of the three poor household models, the random forest model and logistic model are slightly reduced, which are 82% and 72%, respectively, but the waterfall model is basically unchanged, which is 83%, and the three models have little change. The new waterfall design proposed in this article has the advantage of a high percentage of sample reuse and can effectively prevent overfitting, and there is no need for massive data. It is a stable and reliable new model. The combination of targeted poverty reduction algorithms and big information technology and mining data can get the most common causes more accurate and convincing results. The right rib trunk and rib are often separated from the common cause because of the population.

1. Introduction

In the middle of the 20th century, getting rid of poverty has become the primary problem of governments in the international community and one of the problems that must be solved in the national economic development strategy. The main reason is that the global economic level is depressed. All countries regard economic development as the primary task of national development and the core content of international organizations to solve the problem of poverty alleviation [1]. Then, in the 1980s, the international community began to gradually realize that the reason for poverty is not the lack of economic income, but a social problem involving social resources, national policies, medical education, and other aspects, even related to race. China began to move towards the road of not relying solely on economic growth to solve the problem of poverty alleviation, in which sustainable development is one of the main strategies to solve the problem of poverty [2]. From the perspective of the state for the majority of poor households who need to be supported, the previous extensive poverty alleviation model has long been inapplicable to the poverty alleviation population under the current conditions. Afterwards, General Secretary Xi proposed targeted poverty alleviation. Targeted poverty alleviation is no longer the extensive regional poverty alleviation model but targeted at the poor population, that is, the smallest unit in poverty alleviation. The purpose of targeted poverty alleviation is to offset the economic that is one of the necessary measures to be taken to reduce the effectiveness of poverty alleviation due to growth [3, 4].

Poverty has always been one of the major problems faced. In order to ensure that China can build a well-off society as soon as possible, we must solve the problem of poverty from the root [5]. The first step to solve the problem of poverty is to find out the real poor households and eliminate poverty from the source through targeted assistance to the real poor households. As an important part of big data technology, mining algorithm is mainly responsible for mining the potential logical relationship between data. The application core of data mining algorithm is to design the data mining model, use the data mining model to calculate the data, and then mine the potential associations. We need to collect all the data required by the data mining model as the basic data to provide data support for data mining. Then, the type of data is determined as the pattern division standard of data mining model, and the framework of data mining model is determined. Finally, the data mining model is established by summarizing the data characteristic parameters. Although detailed identification studies are not yet fully developed in China, in recent years, with the rapid development of Internet technology and big data technology, it may be faster and more efficient to use large data technology and data mining algorithms to accurately reduce poverty and more accurately identify real poor households. Compared to traditional machine learning algorithms, due to technological limitations and independent storage, they are used only for small amounts of data and are based on data sampling. The advent of big data technology can help us run machine learning activities, modeling on large amounts of data [6, 7]. Figure 1 shows the in-depth integration of big data and targeted poverty reduction.

2. Literature Review

Since the proposal of targeted poverty alleviation has only been put forward for four or five years, and the time in practical application is shorter, there is less research on targeted poverty alleviation. Targeted poverty reduction has been achieved in a short period of time, but with the strong support of the central government and the joint efforts of local governments, targeted poverty reduction has achieved remarkable results in recent years. Today, many Chinese scholars are making creative proposals to reduce poverty and make a detailed description and analysis in combination with the difficulties encountered in practical work. From the perspective of family structure, Zhao S. and others used the poverty measurement method to study the Multidimensional Poverty Situation of different types of family poverty. Finally, they came to the conclusion of what kind of family is more likely to fall into poverty [8]. Awajan and others take advantage of big data knowledge to help targeted poverty alleviation. They compare big data poverty alleviation with traditional poverty alleviation methods and conclude that big data targeted poverty alleviation has more advantages than traditional poverty alleviation methods, which is more conducive to identifying and helping poor households [9]. By Gaye and others traditional poverty alleviation technologies and models have been challenged. In the past, large-scale poverty alleviation models have been less regional-oriented, making it difficult to identify the poorest households. At the same time, many new problems will appear if the data of poor households are not updated for a long time; there is an urgent need to change the traditional model of poverty reduction. The new model of poverty reduction should be the comprehensive use of big data technology, to reduce target poverty and increase the efficiency of targeted poverty reduction [10]. Cui and others believe that there are still some limitations in simply taking the economic income of poor households as the poverty standard of poor households. Therefore, based on the third-party assessment and research task of targeted poverty alleviation, it is found that 13.5% of the surveyed farmers believe that some poor households who really need government assistance are missed by the government in the process of filing and card [11]. Vachkova and others analyzed and defined what targeted poverty alleviation is and expounded their views on why targeted poverty alleviation should be carried out. This paper gives a complete description of the poverty alleviation model since the reform and opening up, analyzes that the previous extensive poverty alleviation model is no longer applicable to the current environment, summarizes the key and difficult points of targeted poverty alleviation, and expounds the new working methods [12]. Visuwasam and others discussed and studied the mechanism of targeted poverty alleviation through big data technology and concluded that big data technology can be widely used in targeted poverty alleviation [13]. Granat and others proposed the research on the rural targeted poverty alleviation mechanism in Guizhou Province under the background of big data. Accurate identification is the most important link in targeted poverty alleviation. How to find out the poor households is the top priority [14]. Srivani and others studied the research on accurate identification in targeted poverty alleviation. The most important thing in exploring accurate identification is to explore the research on accurate identification through Multidimensional Poverty Measurement [15].

This article mainly uses big data technology and data mining algorithms to accurately identify poor households, establishes a variety of different data mining models under the big data computing framework spark, uses the models to identify and classify poor households, and finally compares and evaluates the results of the models. Then it classifies 100 real data with three poverty identification models, checks the identification accuracy of the three models on the real poverty data, and analyzes the modeling time of the three models.

3. Research Methods

3.1. Big Data Analysis Technology. Big data analysis technology: Big data analysis is based primarily on machine learning, cloud computing technology, and several data mining algorithms, but also on some new information analysis technologies. At the same time, it also depends on some new data analysis technologies, mainly including graph-based mining algorithm, group-based mining algorithm, and data network-based mining algorithm. At the same time, it adopts the fusion technology of object-based data connection. In recent years, big data analysis technologies for various fields have emerged one after another,

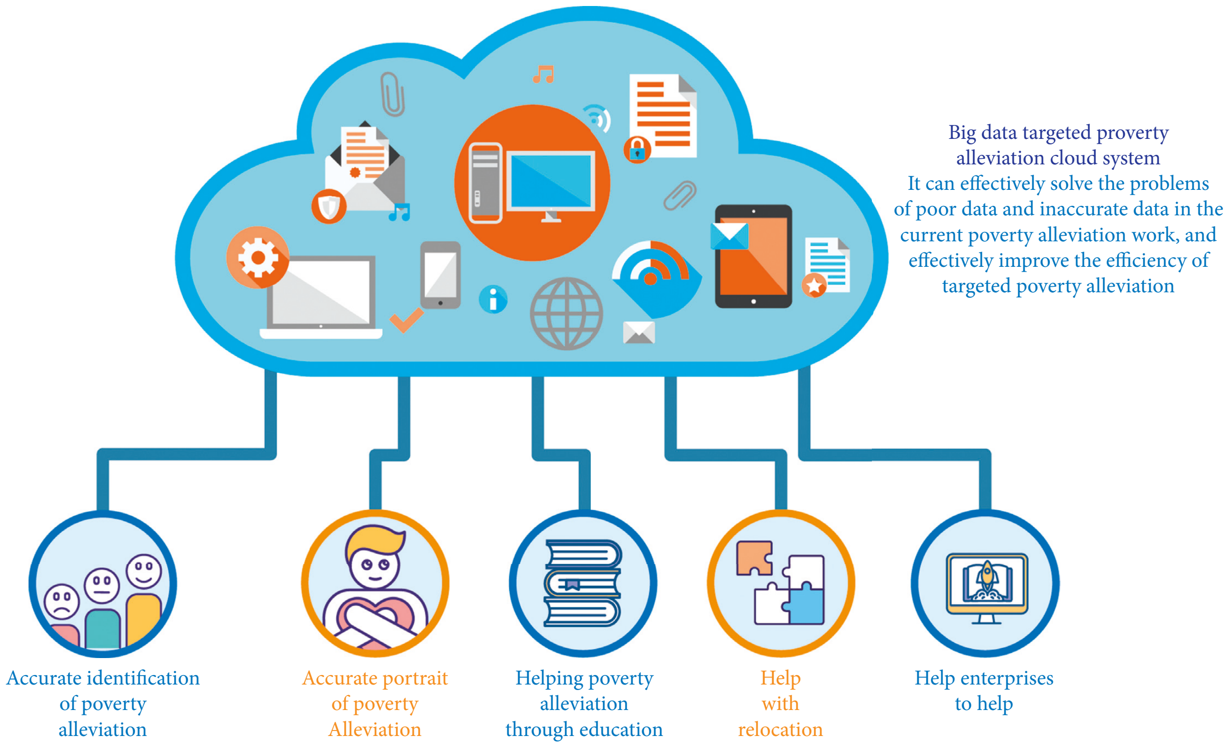


Figure 1: Deep integration of big data and targeted poverty alleviation.

mainly using network behavior analysis, semantic analysis, emotion analysis, syntax analysis, and user interest mining [16].

In practical application, most of the data are data without any rules. The amount of these data is usually very large, but they cannot be stored through relational database. Most of the data are incomplete and have a large amount of noisy data. However, this part of the data hides a large amount of valuable user data, which has very important potential value and academic knowledge. To obtain these valuable data, we must clean these data through data mining algorithms, filter out useless data, and finally retain valuable potential data. The more common data mining methods include clustering, anomaly and trend discovery, association rules, dependency models, etc. The data objects of data mining mainly come from the current mainstream relational databases, text data, web data sources, heritage databases, etc. After data cleaning, it enters the statistical stage. Statistical algorithms mainly include regression analysis and cluster analysis based on data multidisciplinary regression and autoregressive algorithms. Data mining is also known as knowledge discovery in a database. As the name suggests, data mining is from massive and complex data. The process of using certain algorithmic research means to find effective information with certain potential value and revealing significance hidden in data. As a process of decision support, data mining can automatically and intelligently analyze data by using artificial intelligence technology, statistical principle, machine learning, and pattern recognition methods, realize data visualization through visualization technology, make inductive and summary reasoning operations, and mine the laws contained in the data and provide correct guidance for decision makers in formulating strategies, reducing risks and preventing major mistakes [17]. As a multistep processing process, the knowledge in the database includes the operation stages of data selection, preprocessing, data conversion, data mining, result interpretation, and evaluation (see Figure 2).

3.2. Pretreatment of Multidimensional Poverty Characteristics of Farmers

3.2.1. Overview of Maslow's Hierarchy of Needs Theory. Maslow's demand hierarchy theory is one of the theories of behavioral science. According to the pyramid model, human needs from the bottom of the pyramid are divided into five levels: physiological needs, security needs, love and belonging needs, respect, and self-realization needs (Figure 3).

3.2.2. Selection of Poverty Characteristics Based on Maslow's Demand Level. The measurement indicators selected for the multidimensional poverty characteristics mainly form the following at the current stage of development: the basic needs of farmers and the overall characteristics of poor groups, the correlation between research indicators and poverty, the existing poverty research framework, etc. In the selection process, there is not only a lack of corresponding guiding theory, but also many restrictions on the acquisition of data due to various reasons. Therefore, it is difficult to make corresponding trade-offs and lack of norms and systems for the measurement and selection of multidimensional poverty characteristics.

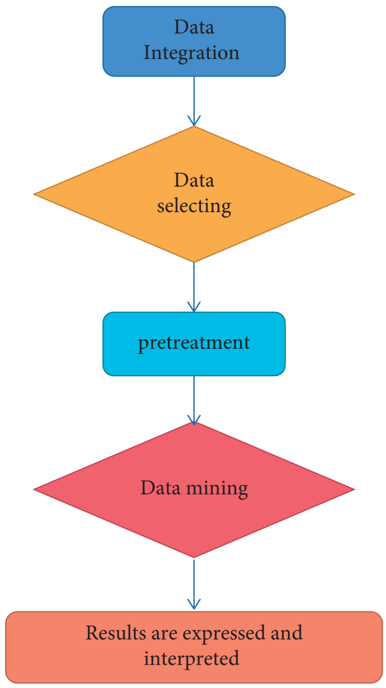


Figure 2: Process diagram of knowledge mining.

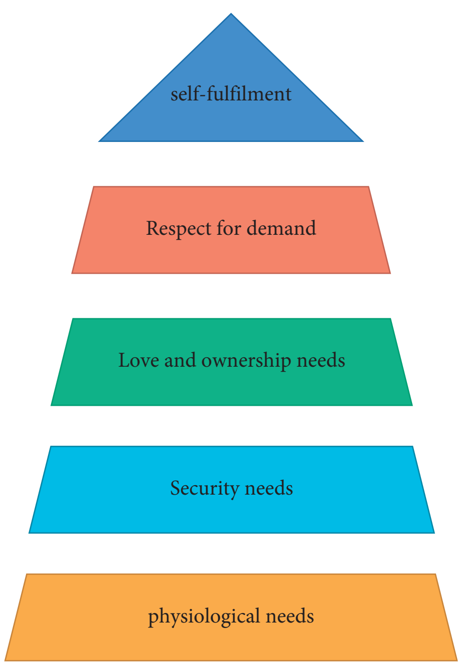


Figure 3: Maslow demand hierarchy pyramid model.

In the targeted poverty alleviation big data platform, various data of farmers are recorded. In addition to farmers' personal information, it also includes the most basic information of various types of farmers, such as food, clothing, housing, and transportation, as well as relatively high-level data on farmers' income, subsidies, medical treatment, and education. As we all know, the impact of various dimensions on the causes of poverty is different. For example, the importance of factors such as food, housing, and transportation must be greater than that of family labor force and education. Only after meeting the conditions of farmers' food, clothing, housing, and transportation can farmers have a higher level of pursuit. Moreover, in the big data platform for targeted poverty alleviation, there are some redundant records of farmers' basic information unrelated to the research objectives, which will affect the results of the experiment. Here, the basic information recorded by farmers in the platform is screened and divided, as shown in Figure 4.

3.2.3. Unsupervised Discretization of Basic Household Information. For the convenience of data mining, for the continuous data contained in the basic information of farmers, the unsupervised discretization of static attributes needs to be carried out before using association rule analysis.

Uncontrolled sorting of farm information means adjusting the interval values of the relevant digital attributes stored in the database and using separate intervals to represent all continuous data, which drastically reduces the number of attribute values and simplifies the initial data. The mining results can be improved with more buttons. The easier-to-operate, more logical representation, equally wide interval method is used for unsupervised discretization of farmer data and differentiation between two continuous data: household income per farmer age and heavy network. For dividing the age structure (Table 1).

Since the national poverty recognition standard for farmers' income is that the per capita annual net income is below 2300, the discretization of per capita net income of farmers' families is formulated according to the national poverty line standard (see Table 2).

The evaluation scores are discretized according to the hundred mark system, with excellent, good, medium, and poor, as shown in Table 3.

3.3. Identification Model of Random Forest Poor Households

3.3.1. Random Forest Algorithm. Random forest algorithm is a new and highly flexible data mining algorithm, an algorithm for integrating multiple decision trees. Therefore, this integrated learning method is often better than the prediction made by any single classification. The basic idea of random forest mainly lies in "random" and "forest." Each time the decision tree is established, some characteristic variables are randomly selected to build the tree through the selected characteristic variables, and then the tree building steps are repeated until the number of decision trees needs to be established. In this way, many independent trees will be established. These trees are the "forest" in the random forest, and the final classification result is determined by the results obtained by the established decision tree through voting [18, 19].

In the application scenario of random forest, it is generally used for classification prediction. The principle of solving the classification problem is usually as follows: generally, when using random forest algorithm for classification, because the random forest is constructed by many decision trees, in order to obtain the final result of the

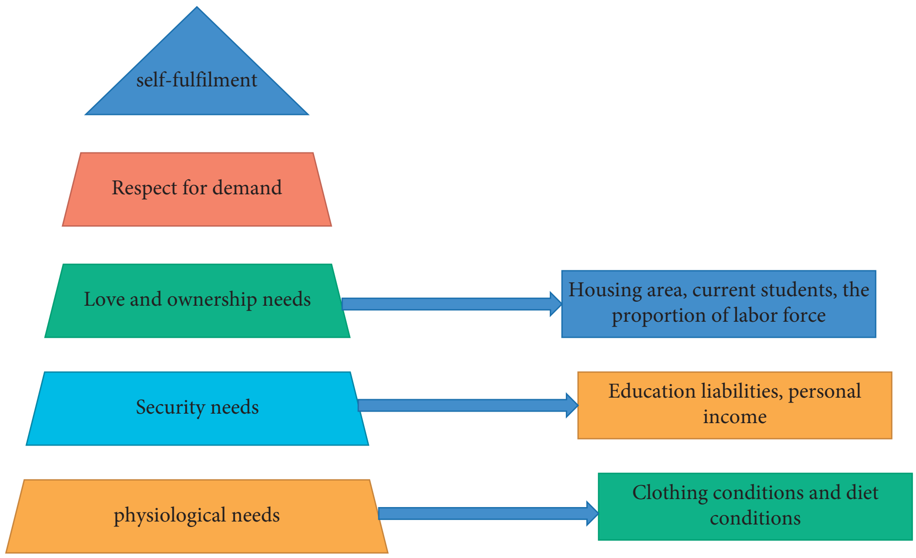


Figure 4: Screening and division of farmers' information.

Table 1: Age discretization structure.

Number	Age	Discretization
1	30 and below	Youth
2	30–49	Mid-life
3	50–59	Quinquagenarian
4	60 and above	Old age

Table 2: Income discretization structure.

Number	Per capita family net income is discrete
1	Under 2300
2	2300–4999
3	5000 and above

Table 3: Evaluation of separation and dispersion structure.

Number	Assess the score	Evaluation scores were discretized
1	More than 89	Ample
2	80–89	Good
3	60–79	Center
4	Less than 60	Difference

random forest classification, it is necessary to vote and summarize the classification results for each decision tree erected. Each decision tree is based on a relatively independent random sample with the same distribution. The creator of random forest combines bagging with random extraction method. This operation not only effectively reduces the correlation between each decision tree, but also improves the performance of random forest tree.

3.3.2. Model Construction. The information in this article is taken from detailed file and card information on poverty reduction in the province, but in order to make the analysis results true and effective, the data also includes noncard filing data. According to the card filing data of poor households and nonpoor households, a total of 34 million pieces of data are generated in proportion, including the data of poor households and nonpoor households. Different from the traditional statistical methods, this paper adopts large data massive sample data set to replace the traditional hierarchical random sampling to construct the data set. Based on both practical and technical considerations, this paper mainly refers to the international multidimensional poverty index in the selection of characteristic variables, which reflects which poverty attributes are easy to make people fall into poverty.

Spark, the big data computing framework used in this paper, is used to model and analyze poor households. First, spark needs to be started and the packages needed to build the model need to be imported. Next, import the original sample data set and preprocess all the data, including feature transformation, which helps to speed up the iterative calculation and model modeling. Finally, the processed data and the random forest model are jointly established to establish pipeline. In the process of random forest modeling, the parameters of the model can be set by setter or paramMap. In this paper, paramMap is selected to set the parameters, and the obtained model parameters are the best model parameters obtained by the final cross validation. Then, by importing the training set data for model training, the random forest poor household identification model, pipeline model, can be obtained, and the pipeline model is used to classify and predict the test set data [20]. The model prediction results are derived together with the real results of the test set and the probability value that the model prediction is a poor household for analysis and comparison. The original sample data used in this paper is 34 million family data. Each family selects the ten characteristic variables described above, and the sample data size is 10.36. According to the above steps, the sample size of the test package, which randomly divided the training package and the test package between 70% and 30%, is about 10.2 million. The results of the random forest model classification on the experimental data are shown in Table 4, and the random forest model evaluation index is shown in Table 5.

The total running time of the random forest poor household identification model under spark is 3 hours and 36 minutes. In the confusion matrix, 1 represents poor households and 0 represents nonpoor households. The overall accuracy of the final result is 89.48% compared to the actual results of the test package, the precision is 91.65%, the recall is 86.68%, the FPR is 0.077, the specificity is 93.23%, and the AUC value is 0.9718. The sum of the number of poor households and nonpoor households identified by the model accounts for 89.48% of the total sample size, 86.68% of all real poor households are identified, and 91.65% of all identified poor households are real poor households. The ROC curve of the model is shown in Figure 5.

Table 4: Results of classification of poor households by the random forest model.

Confusion matrix		Predicted value	
		0	1
True value	0	4631822	387851
	1	662031	4270040

Table 5: Model evaluation indicators.

Model test index	
Overall model identification accuracy	88.37%
Precision ratio	90.54%
Recall ratio	85.57%
Specificity	91.22%
FPR	7.6%
AUC	96.07%

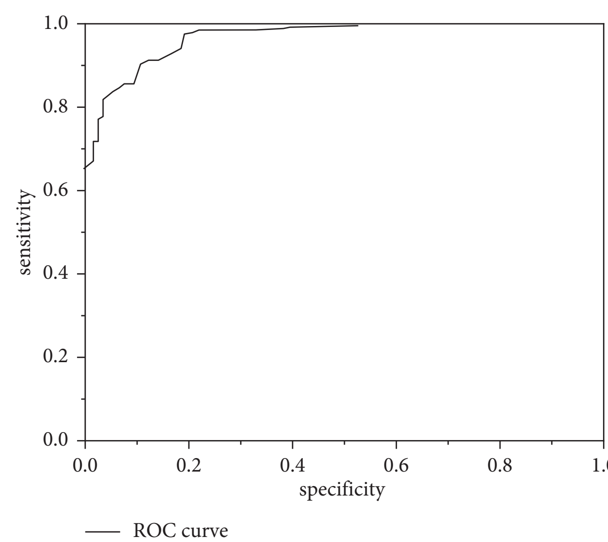


Figure 5: ROC curve of the model.

3.4. Logistics Poverty Identification Model

3.4.1. Introduction to the Logistic Algorithm. Assuming that there is a random variable X, the distribution form of the function of the random variable X is given by equation (1), and the probability density function of the random variable X is given by equation (2).

$$\begin{aligned} F(s) &= P(X \leq s) \\ &= \frac{1}{1 + e^{-(x-u)/r}}, \end{aligned} \tag{1}$$

$$\begin{aligned} f(s) &= F'(s) \\ &= \frac{e^{-(x-u)/u}}{\gamma\left(1 + e^{-(x-u)/r}\right)^2}. \end{aligned} \tag{2}$$

Then it can be considered that x obeys the logistic distribution; in the above formula $\lambda > 0$ can be called shape parameter and u can be called position parameter. Logistic $F(s)$ distribution function is a sigmoid function with different properties. The shape of the sigmoidal function is a curve similar to the s-shape, and the center of symmetry of the function is (μ, 1/2), so it satisfies

$$F(-s+\mu) - \frac{1}{2} = -F(s+u) + \frac{1}{2}. \tag{3}$$

Because the sigmoidal function is an S-shaped curve, the function image grows around the point of central symmetry and changes significantly and changes very little at both ends of the curve. In the distribution function according to formula (3), if the value of the shape parameter Y is smaller, the growth range of the function image S-shaped curve around the center point will be larger [21].

3.4.2. Binomial Logistic Regression Model. Although the binomial logistic regression model is called the regression model, it is actually a classification model, which can be classified according to different segmentation points. The binomial logistic regression model can be expressed by conditional distribution $P(Y|X)$ in the form of parametric logistic distribution. In binomial logistic regression, variable y can be taken as 1 or 0, where x is still a random variable. The conditional distribution of the binomial logistic regression model is shown in

$$P(Y = 1|X) = \frac{\exp(w \cdot x + b)}{1 + \exp(w \cdot x + b)}, \tag{4}$$

$$P(Y = 0|X) = \frac{1}{1 + \exp(w \cdot x + b)}, \tag{5}$$

where $x \in \wp^n$ represents the input vector, $Y \in \{0, 1\}$ represents the predicted output variable, and $x \in \wp^n$ and $b \in \wp^n$ are the parameters of the two models. W is the model weight vector to be estimated by the algorithm, and B is also the parameter to be estimated by the algorithm, representing the offset of the function. Then, compare the probability of 1 with that of 0, take the maximum value, and divide this classification into the category with larger probability value.

In order to facilitate observation and calculation, the model can be simplified by expanding the weight vector to $w = (w^{w(1)}, w^{w(2)}, \ldots, w^{w(n)}, b)^T$ and the input vector to $w = (w^{w(1)}, w^{w(2)}, \ldots, w^{w(n)}, 1)^T$. The internal product of the vectors of this generation is the same as the first, and a simplified logistic regression model is shown in

$$P(Y = 1|X) = \frac{\exp(w \cdot x)}{1 + \exp(w \cdot x)}, \tag{6}$$

$$P(Y = 0|X) = \frac{1}{1 + \exp(w \cdot x)}. \tag{7}$$

The characteristics of logistic regression model: at this time, it is necessary to introduce a new concept, odds ratio odd. Assuming that the probability of an event is p, the probability of its nonoccurrence is $1 - p$. Odds ratio odd refers to the ratio between occurrence and nonoccurrence probability, which is $p/1 - p$. At this time, the logit function of log odds of the event is shown in the following equation:

$$\text{logit}(p) = \log\left(\frac{p}{1 - p}\right). \tag{8}$$

For logistic regression, equation (9) can be obtained by

$$\log \frac{P(Y = 1|X)}{1 - P(Y = 1|X)} = w \cdot x. \tag{9}$$

In the logistic regression model, the odds ratio (odds) of the probability of an event is expressed by the linear function of the input vector x. At the same time, through the definition of logistic regression model in the above two formulas, based on the linear function, the logistic regression model can be transformed into a probability calculation formula (see equation (10)).

$$P(Y = 1|X) = \frac{\exp(w \cdot x)}{1 + \exp(w \cdot x)}. \tag{10}$$

At this point, if you want the probability value of P to be infinitely close to 0, the value of the linear function must be infinitely close to negative infinity. If you want the probability value of P to be infinitely close to 1, the value of linear function should be infinitely close to positive infinity. The final model is called logistic regression model.

3.4.3. Estimation of Model Parameters. Finally, the logistic regression model can be obtained, as shown in equations (11) and (12).

Hypothesis:

$$P(Y = 1|X) = \pi(X), \tag{11}$$

$$P(Y = 0|X) = 1 - \pi(X). \tag{12}$$

Probability function (13) is shown in the following equation:

$$\prod_{I=1}^{N} [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i}. \tag{13}$$

See equation (14) for log likelihood function is as follows:

$$L(w) = [y_i(w \cdot x_i) - \log(1 + \exp(w \cdot x))]. \tag{14}$$

Assuming that w is calculated with the maximum probability of w, then the probability formulas for the true logistic regression model can be obtained by controlled study, as shown in

$$P(Y = 1|X) = \frac{\exp(\widehat{w} \cdot x)}{1 + \exp(\widehat{w} \cdot x)}, \tag{15}$$

$$P(Y = 0|X) = \frac{1}{1 + \exp(\widehat{w} \cdot x)}. \tag{16}$$

3.5. Waterfall Model for Poor Households

3.5.1. Waterfall Design. A waterfall model can be used to make classification assumptions. This model is mainly inspired by the normal distribution graph created by the rapid sand experiment and uses the graph model of the triangular structure (Figure 6).

The middle node of layer 0 accounts for 100% of the total data, while the middle node of layer 2 accounts for 50% of the total data, and the middle node of layer 4 accounts for 37.5% of the total data. By analogy, the proportion of the middle node of layer $2n$ in the total data is $C2n$. This is because by Stirling formulas:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{(1/12n)-(1n(9n)/(9n)^{\pi}-(9n)^{-\pi})}, \tag{17}$$

$$(n!)^2 \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^{2n} e^{(1/6n)-(21n(9n)/(9n)^{\pi}-(9n)^{-\pi})}, \tag{18}$$

$$(2n!) \approx \sqrt{2\pi 2n}\left(\frac{2n}{e}\right)^{2n} e^{1/24n-(1n(18n)/(18n)^{\pi}-(18n)^{-\pi})}. \tag{19}$$

Therefore, when $n \longrightarrow \infty$, the proportion of intermediate nodes in the total data tends to be close to 0; that is, the limit is 0.

3.5.2. Construction of the Waterfall Model. Due to the waterfall, it is more convenient to select the parameters of the model [22, 23]. First, start the big data computing framework spark. Since the waterfall model is a new model; there is no model in spark's machine learning library pipeline. When importing the required package, you only need to import other packages required for building the model. The sample data is also 34 million pieces of data, with a size of 10.36. The results of the waterfall model classification and assumptions in the experimental set of data are shown as follows. The actual class label in the left test set represents the waterfall model, which predicts the class label of the test set data and the predicted value on the right. The evaluation parameters of the waterfall model are shown in Tables 6 and 7.

The overall running time of spark is only 1 hour and 42 minutes. The actual results of the test package were

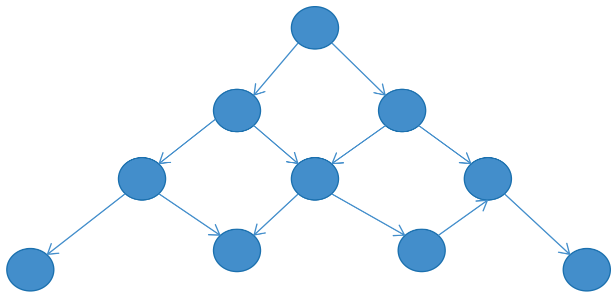


Figure 6: Waterfall model.

Table 6: Results of waterfall model classification of poor households.

Confusion matrix		Predicted value	
		0	1
True value	0	4452260	561631
	1	958013	4077148

Table 7: Model evaluation indicators.

Model test index	
Overall model identification accuracy	83.71%
Precision ratio	86.54%
Recall ratio	80.72%
Specificity	87.70%
FPR	10.1%
AUC	92.34%

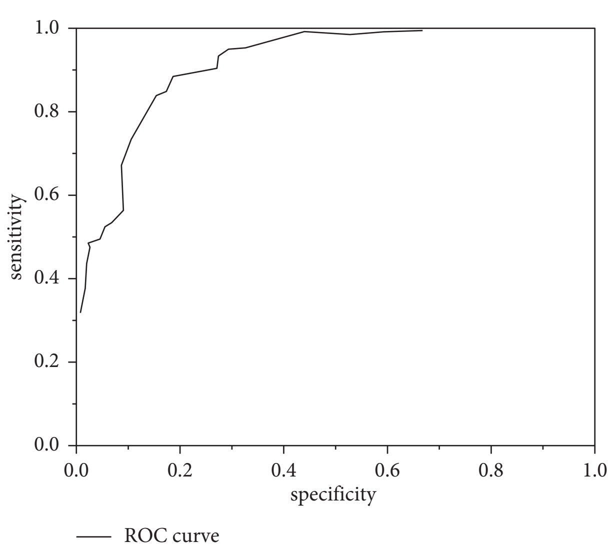


Figure 7: ROC curve of the model.

compared with the results predicted by the model, and the overall identification accuracy of the model was 83.71%, the accuracy was 86.54%, the recall rate was also the recall rate, and the TPR was 80.72%. . The feature is 87,700%, the FPR value is 0.1120, and the AUC value of the area along the ROC curve is 0.9234. The model is shown in the ROC curve in Figure 7.

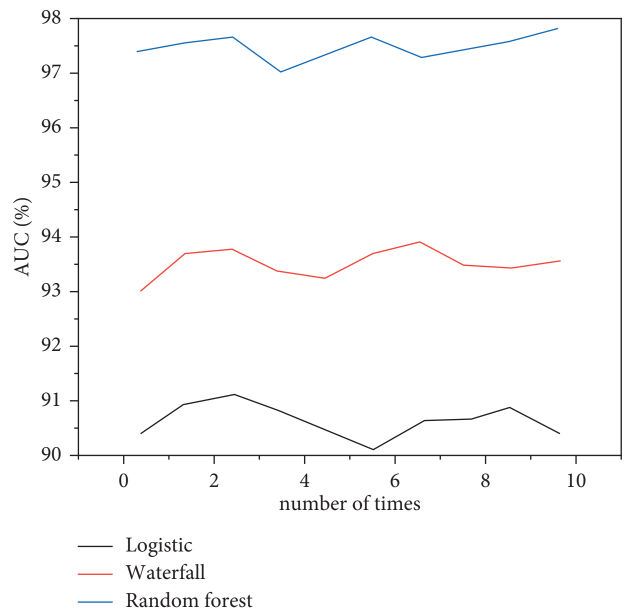


Figure 8: 10 AUC values of three models.

4. Results and Discussion

4.1. Multiple Selection Test Set Evaluation Model. In order to compare and evaluate the three models for identifying poor households, a new set of data that was not included in the model training and experiment was selected from the three sets of pilot models. The data with a data size of 30% of the total sample size are randomly selected from the new test set for times as the test set; that is, the data size of each randomly selected test set is about 10.2 million poverty-stricken households. Three models were tested to compare the average AUC of the three models and the stability of the three models [24]. In 10 experiments, the AUC of the random forest model was consistently higher than that of the logistics model and the waterfall model. The average AUC was 0.934 (Figure 8).

The changes of AUC values of the three models are small, which proves that the stability of the three models is very good. However, the operating time is much shorter than the other two models, with a random forest model running time of 3 hours and 36 minutes, a logistics model running time of 2 hours and 20 minutes, and a waterfall running time of only 1 hour and 42 minutes. It is more than half an hour of running even a small random forest model.

4.2. Testing the Model with Real Data. This is because 34 million pieces of data are generated proportionally from the filing and card filing data of poor households when building the model [25]. In order to test the accuracy of the constructed model on the filing and card filing data of fully real poor households, 100 fully real poor households' data were randomly selected, including 50 real poor households and 50 nonpoor households, and 100 real data were used to test the

Table 8: Results of the classification of poor households by the random forest model.

Confusion matrix		Predicted value	
		0	1
True value	0	41	7
	1	8	40

Table 9: Classification results of poor households by the logistic model.

Confusion matrix		Predicted value	
		0	1
True value	0	36	12
	1	13	35

Table 10: Results of waterfall model classification of poor households.

Confusion matrix		Predicted value	
		0	1
True value	0	41	7
	1	7	41

three models, respectively. The established random forest poverty identification model is used to test 100 real data. For the test results, see Table 8.

According to the classification results of 100 real data by the random forest poor households identification model constructed by spark, the overall identification accuracy of the model is 83%, and 41 of the 50 poor households are identified. 42 out of 50 nonpoor households were identified. The recognition accuracy of random forest poor households recognition model for 100 real data is slightly reduced.

The constructed logistic poor household identification model is used to test 100 real data, and the test results are in Table 9.

Through the classification results of 100 real data by the logistic poor households identification model built by spark, a big data computing framework, it can be seen that the overall identification accuracy of the model is 72, and 36 of the 50 poor households have been identified. The accuracy of the logistic poor household identification model in identifying the real data of 100 households also decreased slightly.

The constructed waterfall poor household identification model is used to test 100 real data, and the test results are in Table 10.

According to the classification results of 100 real data by the waterfall poor households identification model constructed by spark, the big data calculation framework, the overall identification accuracy of the model is 84%; 42 of the 50 poor households and 42 of the 50 nonpoor households are identified. The recognition accuracy of waterfall poor household recognition model for 100 real data is basically unchanged. When 100 real data are used to test the accuracy of the three poor household models, the random forest model and logistic model are slightly reduced, which are 82% and 72%, respectively, but the waterfall model is basically unchanged, which is 83%, and the three models have little change.

5. Conclusion

The essence of targeted poverty alleviation is that the government effectively identifies poor families and members, excavates the causes and extent of poverty, and carries out practical and effective assistance, so as to fundamentally break the barriers of poverty. With the rapid economic development, the national income level is seriously unbalanced. The previous extensive regional poverty alleviation method has long been inapplicable. In this case, targeted poverty alleviation came into being. Through the research, the logistic algorithm, random forest algorithm, and the newly proposed waterfall model in data mining are found. The newly proposed waterfall model has the advantages of high sample reuse rate, can effectively prevent overfitting, and has no demand for massive data. It is a stable and reliable new model. Visualize the research results, design and implement the abnormal poverty assistance model obtained in the above research work, and apply it to the actual poverty alleviation work, so that the staff can get the effectiveness of the government's policy assistance to the village by importing the basic information and assistance data of farmers into the system. It can make timely and effective adjustments to the existing abnormal assistance phenomenon, so as to provide help for the government in the targeted poverty alleviation work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was sponsored by Jiaozuo Normal College.

References

[1] H. Yao and J. Fu, "Analysis of smooth implementation of industry poverty alleviation considering government supervision," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5554595, 13 pages, 2021.

[2] R. Narayan, S. Gadag, S. P. Cheruku et al., "Chitosan-glucuronic acid conjugate coated mesoporous silica nanoparticles: a smart ph-responsive and receptor-targeted system for colorectal cancer therapy," *Carbohydrate Polymers*, vol. 261, no. 2, pp. 117893–117898, 2021.

[3] J. Wu, Z. Feng, and S. Anwar, "Assessing the pollution convergence across Chinese cities by considering ecological indicators: a continuous distribution dynamics approach," *Ecological Indicators*, vol. 126, no. 2, pp. 107552–107559, 2021.

[4] B. He and L. Yin, "Prediction modelling of cold chain logistics demand based on data mining algorithm," *Mathematical Problems in Engineering*, vol. 2021, Article ID 3421478, 9 pages, 2021.

[5] B. Khan, R. Naseem, M. A. Shah et al., "Software defect prediction for healthcare big data: an empirical evaluation of machine learning techniques," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8899263, 16 pages, 2021.

[6] D. Wang, T. Miwa, and T. Morikawa, "Big trajectory data mining: a survey of methods, applications, and services," *Sensors*, vol. 20, no. 16, pp. 4571–4577, 2020.

[7] D. Xia, F. Ning, and W. He, "Research on parallel adaptive canopy-k-means clustering algorithm for big data mining based on cloud platform," *Journal of Grid Computing*, vol. 18, no. 2, pp. 263–273, 2020.

[8] S. Zhao, M. Hu, Z. Cai, Z. Zhang, T. Zhou, and F. Liu, "Enhancing Chinese character representation with lattice-aligned attention," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2021.

[9] I. Awajan, M. Mohamad, and A. Al-Quran, "Sentiment analysis technique and neutrosophic set theory for mining and ranking big data from online reviews," *IEEE Access*, vol. 9, no. 99, pp. 1–16, 2021.

[10] B. Gaye, D. Zhang, and A. Wulamu, "Improvement of support vector machine algorithm in big data background," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5594899, 9 pages, 2021.

[11] Y. Cui, "Intelligent recommendation system based on mathematical modeling in personalized data mining," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6672036, 11 pages, 2021.

[12] S. N. Vachkova, E. Y. Petryaeva, R. B. Kupriyanov, and R. S. Suleymanov, "School in digital age: how big data help to transform the curriculum," *Information*, vol. 12, no. 1, pp. 33–36, 2021.

[13] L. M. M. Visuwasam and D. P. Raj, "A distributed intelligent mobile application for analyzing travel big data analytics," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2036–2052, 2020.

[14] J. Granat, J. M. Batalla, C. X. Mavromoustakis, and G. Mastorakis, "Big data analytics for event detection in the iot-multicriteria approach," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4418–4430, 2020.

[15] B. Srivani, N. Sandhya, and B. Padmaja Rani, "Literature review and analysis on big data stream classification techniques," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 24, no. 3, pp. 205–215, 2020.

[16] J. Liu, G. Shi, J. Zhou, and Q. Yao, "Prediction of college students' psychological crisis based on data mining," *Mobile Information Systems*, vol. 2021, Article ID 9979770, 7 pages, 2021.

[17] Y.-X. Shi, B.-K. Zhang, Y.-X. Wang, H.-Q. Luo, and X. Li, "Constructing crop portraits based on graph databases is essential to agricultural data mining," *Information*, vol. 12, no. 6, pp. 227–231, 2021.

[18] X. Wei, "A classification method of tourism English talents based on feature mining and information fusion technology," *Mobile Information Systems*, vol. 2021, Article ID 5520079, 9 pages, 2021.

[19] S. Khan and S. Alqahtani, "Big data application and its impact on education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 17, pp. 36–42, 2020.

[20] S. Zhao, M. Hu, Z. Cai, and F. Liu, "Dynamic modeling cross-modal interactions in two-phase prediction for entity-relation extraction," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2021.

[21] T. Ramalingeswara Rao, S. K. Ghosh, and A. Goswami, "Mining user-user communities for a weighted bipartite network using spark GraphFrames and Flink Gelly," *The Journal of Supercomputing*, vol. 77, no. 6, pp. 5984–6035, 2021.

[22] E. Prasetyo and A. Putra, "Implementasi waterfall model dalam pengembangan sistem informasi eksekutif penduduk," *Journal of Information Systems and Informatics*, vol. 3, no. 1, pp. 213–224, 2021.

[23] N. Rofiq, A. Perdananto, and N. Jaya, "Penerapan model waterfall pada aplikasi bank sampah," *Infotech Journal of Technology Information*, vol. 7, no. 1, pp. 19–26, 2021.

[24] D. J. Kalita, V. P. Singh, and V. Kumar, "Two way threshold based intelligent water drops feature selection algorithm for accurate detection of breast cancer," *Soft Computing*, vol. 26, no. 5, pp. 2277–2305, 2021.

[25] G. A. Loosmore and R. T. Cederwall, "Precipitation scavenging of atmospheric aerosols for emergency response applications: testing an updated model with new real-time data," *Atmospheric Environment*, vol. 38, no. 7, pp. 993–1003, 2004.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9826897, 1 page
https://doi.org/10.1155/2023/9826897

Retraction

Retracted: Evaluation System of Foreign Language Teaching Quality Based on Spatiotemporal Feature Fusion

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Hao and Y. Zhu, "Evaluation System of Foreign Language Teaching Quality Based on Spatiotemporal Feature Fusion," *Security and Communication Networks*, vol. 2022, Article ID 9390358, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9390358, 9 pages
https://doi.org/10.1155/2022/9390358

Research Article

Evaluation System of Foreign Language Teaching Quality Based on Spatiotemporal Feature Fusion

Yan Hao and Yi Zhu

School of Software, Dalian University of Foreign Languages, Dalian, Liaoning 116044, China

Correspondence should be addressed to Yan Hao; haoyan@dlufl.edu.cn

Received 31 December 2021; Revised 23 January 2022; Accepted 31 March 2022; Published 23 April 2022

Academic Editor: Chin-Ling Chen

How to evaluate the teaching quality of foreign language teachers objectively and quantitatively is one of the important directions of teaching evaluation institutions and teaching and research personnel. To solve this problem, this article proposes a foreign language teaching quality evaluation system based on the integration of spatiotemporal features. Given the behavior characteristics of multiperson interaction in class, a framework-based spatiotemporal modeling method is presented in this article. Spatiotemporal modeling features are input into generalized graph convolution for feature learning. The interaction information between skeletons is designed to capture the extra interaction information to increase the accuracy of action recognition. The experimental results show that the proposed method has higher accuracy and can be applied to the evaluation of foreign language teaching quality.

1. Introduction

The quality of foreign language classroom teaching is mainly determined by two key links, namely, the teacher's teaching ability and the degree of students' receiving knowledge in class. The following problems have become the focus of teaching evaluation institutions and teaching and research personnel: how to evaluate the quality of teaching objectively and quantitatively, how to accurately count the students' understanding and mastery of each knowledge point, and how to intuitively reflect the advantages and disadvantages of different teachers' teaching effects on the same course [1].

In recent years, with the development of artificial intelligence technologies such as deep neural network and ultra-large-scale data feature analysis, the recognition accuracy of computer image recognition, speech recognition, and emotion recognition has been greatly improved. Human-machine language dialogue, AI customer service, AI simultaneous interpretation, and other technologies have been gradually commercialized. In the field of teaching, various virtual experimental teaching and large-scale online classroom supported by AI technology have been applied to practical teaching [2]. These technologies have greatly expanded the way teaching content is presented. At the same time, the application of 3d visual image, virtual real scene, simulation experiment, and other technical means makes the traditional design and experiment process more vivid [3]. These technologies can help schools timely grasp the degree of students' response and recognition to classroom teaching content. In this context, how to use AI technology to help teachers and teaching departments assess the quality of classroom teaching more accurately, objectively, and efficiently has important research value.

Video-based interactive behavior recognition has a high practical value and broad application prospects. The purpose of human motion recognition is to analyze and understand the actions and interactions between people in video. It can be applied in intelligent monitoring, human-computer interaction, video sequence understanding, and medical and health and other fields, playing an increasingly important role in daily life [4].

In behavior recognition, the application of human skeleton data has obvious advantages over RGB data and depth data. It can be unaffected by background, lighting, and appearance. In addition, skeleton features compact, strong structure, and rich semantics. It has strong ability to describe human movements and motions, and more and more behavioral recognition studies are carried out based on skeleton. At present, there are three skeleton interactive recognition methods based on deep learning: long- and short-term memory network, convolutional neural network, and graph convolutional network [5]. For the time being, relatively mature studies are focused on the recognition of single skeleton actions, and there is a lack of discussion on interactive actions. However, in daily life, common behaviors are basically some interactive ones, such as shaking hands, hugging, fighting, and so on. Interactive action is more complex than solo action. In the process of completing interactive movements, there are more types of body movements, and the changes between body movements are also more diversified. Therefore, how to effectively extract the characteristics of interactive actions, and conduct modeling and analysis of interactive behavior is a very challenging problem.

The previous work was to reorganize the skeleton data into a grid structure, which was implemented by RNN (recurrent neural network) and CNN (convolutional neural networks) [6]. Although they have made great improvements in motion recognition, there are still some problems. Because human skeletons are graphical structures, rather than traditional fixed grids, they do not fully benefit from the superior representation capabilities of deep learning. The human skeleton is a naturally constructed figure in a non-Euclidean space. Although the CNN method has strong feature extraction capability, it requires a convolution kernel of fixed size for ergodic processing. Therefore, it cannot extract key features of graph data effectively, and its computational complexity is large. It cannot meet the accuracy requirement in multitask processing, which makes the traditional convolutional neural network not applicable. Although the traditional RNN can also process the skeleton, the accuracy of skeleton data transformation and recognition is not high. Therefore, in this article, GCN is used to process the transformed skeleton data to capture the motion space features [7]. A variant RNN structure is used to capture the temporal dependence information. GCN can directly model the raw skeleton data, extend the graph neural network to the spatiotemporal graph model, and automatically learn spatial and temporal information from the skeleton. The introduction of GCN into skeleton-based motion recognition has yielded many encouraging results. However, most GCN methods are based on predefined graphs with fixed topological constraints, ignoring the implicit joint associations. Meanwhile, GCN could not capture the time information of the whole action sequence completely and could not obtain the action sequence dependence information [8].

To solve this problem, various adaptive connections are designed in this article. It emphasizes the relationship between individuals, interaction objects, and time frames. Meanwhile, feature extraction of time series-dependent information is enhanced. During the recognition of interactive actions, additional information from the interaction itself can be extracted by modeling the interaction relationship between each part of the participant's body. This information is used in global descriptors to identify human interactions to improve the accuracy of interaction action recognition. This article proposes a framework-based spatiotemporal modeling method, which not only designs the connections of a single object and multiple objects in a single frame, but also combines the different connections of single frame and multiple frames. An effective representation of the interactive skeleton diagram is achieved by connecting the relevant joints in the previous frame and the next frame.

The innovations and contributions of this article are listed as follows:

(1) In this article, slice RNN is innovatively applied to the field of video action recognition to enhance the extraction of video sequence-dependent information.
(2) Meanwhile, the spatiotemporal modeling method combined with slice RNN can effectively remedy the disadvantages of slice RNN.
(3) Finally, the algorithm is applied to the foreign language teaching quality assessment system. The experimental results show that the proposed method has higher accuracy and can be applied to the evaluation of foreign language teaching quality.

The structure of this article is listed as follows. Related work is described in the next section. The proposed system is expressed in Section 3. Section 4 focuses on the experiment and analysis. Section 5 is the conclusion.

2. Related Work

2.1. Image-Based Interactive Recognition. Much of the early recognition work was based on manually constructed features, for example, using directional gradient histogram and optical flow directional information histogram [9] to extract appearance features of static information, or using optical flow to extract motion features of dynamic information. Newer approaches rely on deep learning. Literature [10] uses the deep learning network for interactive behavior recognition, extracting optical flow characteristic information through CNN and then feeding it into a classifier to realize action recognition.

Although the motion recognition method based on RGB video or optical flow has high performance, there are still some problems. For example, it is easily affected by background, illumination, and appearance changes, and it requires high computational cost to extract optical flow information. Some work has been done to extract bone data to avoid learning interaction patterns directly from videos. In some research studies on single-person motion recognition, most scholars use human skeleton for motion recognition. The human skeleton can well represent the movement of the human body, which is helpful to analyze it. On the one hand, skeleton data are inherently robust in background noise, providing abstract and high-level features for human motion. On the other hand, skeleton data are very small compared to RGB data. This allows this article to design a better model. Therefore, this article expands skeleton-based motion recognition from single person to multiple people.

2.2. Bone-Based Interactive Recognition. With the development of deep learning, bone-based approaches are emerging. Literature [11] proposes a spatiotemporal LSTM network of node sequences. It extends the learning of LSTM to time domain, and each joint receives information from adjacent joints as well as the previous frame to encode spatiotemporal features. Then, a tree structure is used to represent the adjacent characteristics and motion relations between the nodes. Finally, the results of skeleton data are sent to LSTM network for modeling and identification. Literature [12] divides human skeleton into 5 parts according to the physical structure of a human body and divides them into 5 bidirectional recursive connected subnets, respectively. The researcher proposes an end-to-end spatiotemporal attention model for identifying human actions from skeletal data [13]. Based on LSTM and RNN, a spatial attention module with a joint selection gate is designed. It adaptively allocates different attention to different joints of the input frame within each frame. There are also some methods based on CNN. For example, literature [14] represents bone sequences as a series of enhanced visual and motion color images. The method implicitly describes the spatiotemporal skeletal joints in a compact and unique way. Studies also combine convolutional neural networks with recursive neural networks to perform more complex temporal reasoning for interactions. In view of the good performance of RNN and CNN in skeleton-based action recognition, literature [15] proposed a deep network structure combining CNN classification with RNN. It realizes the attention mechanism of human interaction recognition. The method based on RNN has strong ability of sequence data modeling, and the method based on CNN has good parallelism, and the training process is relatively simple. But neither CNN nor RNN can fully represent the structure of the skeleton. Literature [16] proposed a graph-based regression GCN method for skeleton-based motion recognition to capture spatiotemporal changes in data. However, these methods do not have explicit graph constructs in the actions of identifying interactions. This article further uses the relationship between skeletons to extract interactive features between human bodies. The graph convolution is combined with RNN to better extract the dependency information between nodes and frames.

3. The Proposed Evaluation System of Foreign Language Teaching Quality

3.1. Intraframe Interaction Modeling. The connection of key points is divided into single-person connection within the frame, interactive connection, and interframe connection. These connections are designed by different methods, and then, the spectral convolution is used to obtain the variation characteristics. Then, sequence-dependent information is acquired by combining with slice RNN for action recognition.

In-frame design is divided into single-person design and interactive design. For a single person in each frame, the human body is modeled by a connected graph. The human body connectedness graph is only represented by the natural connection of individuals, so the global information of the human body cannot be well extracted. Therefore, it can be divided into internal connection and external connection through different correlations between nodes. Internal connections include physical connections between joints, and external connections represent potential connections between joints that are not physically connected. There is no skeletal connection between the human hand and head during communication. But because people generally place their hands in front of their bodies, there is an underlying relationship between the hands and the head. Establish external connections between them. Different parameters are set in the weighted adjacency matrix to distinguish the two relations. As shown in Figure 1, the internally connected edge and externally connected edge are given different weights, and the weight of the intraframe edge is set as

$$m_{x,y} = \begin{cases} \alpha, & (x, y) \in \varepsilon_1, \\ \beta, & (x, y) \in \varepsilon_2, \\ 0, & x = y, \end{cases} \tag{1}$$

where $m_{x,y}$ indicates that the joints are not connected. α and β indicate the parameters set for internal and external connection, respectively. In addition, the connection between the joints was represented by ε_1 and ε_2, respectively. ε_1 represents the internal connection between the joints, as shown by the solid black line in Figure 1. As an important property, the distance between the connecting nodes remains constant during motion. ε_2 represents the external connection between the joints, as shown by the dotted line in Figure 1. External dependence refers to the disconnection between two joints, which is also an important factor in the process of motion.

Unlike previous work, in bone interaction recognition, the joints between two people are disconnected. Learning how to describe how each object relates to each other is necessary to merge two people and their interaction information. By analyzing the structure of the bones between two people, information about their interactions can be extracted. Interaction design was carried out between participants of the action, and two independent skeleton diagrams were connected through the joint. Then, they are integrated into an action skeleton diagram with interactive information, and the interactive information of actions can be extracted through the graph convolutional network.

Interaction design consists of two parts, and the joining of points prone to joint-like changes is called correspondence joining. Using ε_3 to represent corresponding associations, such as hugging, the two participants performed roughly the same. Establish connections between the corresponding gateways, as shown in the dotted line in Figure 2. These correspondences play an important role when the participants' actions are generally consistent. In addition, connections between other nodes are called potential connections. The potential connections are indicated by ε_4 as

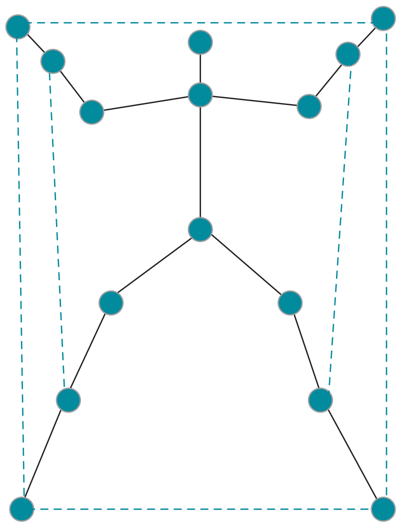


Figure 1: Interframe connectivity diagram.

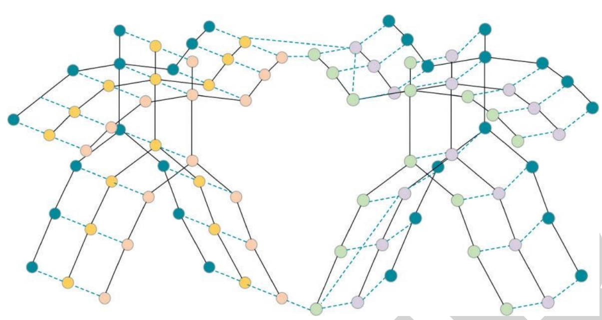


Figure 2: Interframe connectivity of multiple people.

represented by the dotted line in Figure 2. Assign θ to the weight of the edge in ε_3 and δ to the weight of the edge in ε_4, that is,

$$m_{x,y} = \begin{cases} \theta, & (x, y) \in \varepsilon_3, \\ \delta, & (x, y) \in \varepsilon_4, \\ 0, & \text{other}, \end{cases} \tag{2}$$

where x and y represent the key points of different people. The adjacency matrix within a single frame is expressed as

$$G_* = \begin{bmatrix} U_1 & W_1 \\ W_2 & U_2 \end{bmatrix}, \tag{3}$$

where U_1 and U_2 describe the one-person connection. W_1 and W_2 describe the interconnections.

In order to determine which nodes are connected for the interaction modeling within the above frame, the correlation between interaction nodes is measured by Euclidean distance. Calculate the value between all points:

$$d\left(i_x, i_y\right) = i_x - i_y^2, \tag{4}$$

where i_x and i_y are the feature representations of key points x and y, respectively. Compute only the Euclidean distance of the externally connected and potentially connected edges. Then, the resulting distance is normalized. The results are mapped to between [0, 1], and the normalization method of maximum and minimum values is adopted, that is,

$$t = \frac{d_x - d_{\min}}{d_{\max} - d_{\min}}, \tag{5}$$

where $d_{\max}$ represents the maximum joint distance and $d_{\min}$ represents the minimum joint distance.

In this article, a new edge connection is generated when $t < 0.3$ is set experimentally. This not only adds some new necessary interconnections, but also gives the underlying graph some sparsity.

3.2. Interframe Modeling

3.2. Interframe Modeling. Each joint is disconnected in the time domain, allowing each joint in frame i_n to be connected to its corresponding neighborhood in previous frame i_{n-1} and later frame i_{n+1}, as shown in Figure 2.

Extending the receptive field by using more adjacent joints can help the model learn information about the changes in the time domain. These adjacent joints include two types: joints within the same video frame (intraframe joints) and joints between two video frames (interframe joints). The corresponding joint is represented as ε_5. The connected connection between each joint and the neighborhood of the corresponding joint in the adjacent frame is expressed as ε_6. The weights of these two kinds of edges are expressed as

$$m_{x,y} = \begin{cases} \gamma, & (x, y) \in \varepsilon_5, \\ \lambda, & (x, y) \in \varepsilon_6, \\ 0, & \text{other}, \end{cases} \tag{6}$$

where c and y represent the nodes between different frames. The finally constructed multiframe adjacency matrix is expressed as

$$G_{\text{total}} = \begin{bmatrix} G_{*(n-1)} & G_{n-1,n} & 0 \\ G_{n,n-1} & G_{*(n)} & G_{n,n+1} \\ 0 & G_{n+1,n} & G_{*(n+1)} \end{bmatrix}, \tag{7}$$

where $G_{*(n)}$ represents the adjacency matrix of the in-frame modeling graph of frame x. $G_{x,y}$ represents the adjacency matrix between frame x and frame y. Zero is the zero matrix. The calculated graph Laplace is thus $L = D - G_{\text{total}}$.

3.3. Spectral Convolution Algorithm Based on Connected Graph. The skeleton diagram is constructed by taking joints as nodes and the connections between nodes as edges. In a frame, joints are connected internally and externally to act as spatial edges. Interframe connections act as time edges, and the property of each node is the coordinate vector of the joint. The spectral convolution operation is applied to the spatiotemporal skeleton graph to obtain an advanced feature graph.

Consider an undirected graph $A = \{Q, E, G\}$ consisting of vertex set Q and edge set E connecting vertices and weighted adjacency matrix G. G is a real symmetric matrix, and $g_{(xy)}$ is the weight assigned to the edges (x, y) connecting vertices x and y. Assume that the weight is non-negative. Laplacian matrices defined by adjacency matrices can be used to reveal many useful properties of graphs. In different variations of the Laplace matrix, the combinatorial graph used is defined by Laplace:

$$L = D - G, \tag{8}$$

where the Laplace definition of symmetry normalization is $\widetilde{L} = D^{-1/2} L D^{-1/2}$. D is the degree matrix of $d_{xx} = \sum_{y=1}^{t} g_{x,y}$.

The basis of skeleton-based motion recognition is to capture the changes of joints and learn motion features for classification. Use Laplace to simulate the changes in bone. Laplace matrix L is essentially a high-pass operator that can capture the changes of underlying signals. In order to adapt the input sequence length to the input requirements of slice RNN, a full connection layer is used to adjust the data dimension. Finally, the output classification is generated by softmax activation function.

3.4. Timing Sequence Modeling Based on Slice RNN. Interframe modeling has been carried out above to expand the receptive field and learn time-domain change information. However, this kind of interframe modeling cannot capture the time information of the whole action sequence completely and cannot obtain the action sequence dependence information. Therefore, RNN is used in time series processing to solve the dependency problem of action sequence data. However, the current node information of the traditional RNN network is only related to the previous node, so it can only model short-term dynamic information and cannot store long-term sequence. Meanwhile, the standard RNN network structure cannot realize parallel computation like the CNN network model, so the slicing RNN network model is adopted to solve the above problems.

The input sequence is divided into multiple sequence segments, and an independent RNN network is used to calculate each segment. In this article, the RNN hiding unit adopts the gated cyclic unit (GRU), which not only realizes the “parallelism” of computation, but also performs RNN feature extraction on each relatively short sequence fragment. The transfer of information between layers allows for a greater degree of retention of information about long-term dependencies. H represents the hidden layer state of the network, and Y represents the top-level output. The input data itself can compensate for the loss of long-term dependence at the slice point through interframe modeling.

At level 0, the recursive unit acts on each of the smallest sequences by joining structures. Then, the last hidden state of each smallest sequence at level 0 is obtained and used as input to the parent sequence at level 1. The last hidden state of each subsequence at $u-1$ layer is used as the input of its parent sequence at u layer. The last hidden state of the subsequence on the u layer is calculated:

$$b_n^1 = \mathrm{GRU}^0\left(mss^0_{(n-l_0+1)\sim n}\right), \tag{9}$$

$$b_n^{u+1} = \mathrm{GRU}^u\left(b^u_{n-l_u} \sim b^u_n\right), \tag{10}$$

where l_0 represents the smallest sequence length at layer 0. l_u represents the minimum sequence length of u layer. b_n^u represents the hidden layer representation of the n subsequence of u layer. mss^0 represents the smallest sequence at layer 0. $mss^0_{(n-l_0+1)\sim n}$ is the calculation of hidden state in subsequence at layer 0. Different GRUs can be used for different layers. Equation (10) indicates that after the hidden state is calculated at layer 0, the next hidden state is calculated again with the calculation result, and the calculation is repeated. This operation is repeated between each subsequence on each layer until the final hidden state of the top layer (z-th layer) is obtained:

$$F = \mathrm{GRU}^z\left(b^z_{n-l_z} \sim b^z_n\right). \tag{11}$$

Similar to standard RNN, the softmax layer is added after the final hidden state F to classify video actions, that is,

$$u = \mathrm{softmax}\left(M_F F + h_F\right). \tag{12}$$

3.5. Design of Evaluation System. This system is capable of analyzing continuous images and identifying human behavior characteristics from images, which is shown in Figure 3. The system realizes the functions of image interpretation and transcoding, image preprocessing, and face moving optical flow tracking by OpenCV. The algorithm in this article realizes the recognition of human morphological features. Finally, Python and deep learning architecture library are used to realize the recognition of facial features, including the performance of students' specific behaviors such as nodding, bowing, and sleeping.

Supported by the above technologies, this article analyzes the classroom video collected by the camera in real time. At present, it can recognize and output information mainly including the following three points:

(1) Students' attendance. The number of students in a class can be calculated through the recognition of human morphological features. Combined with the information of courses and classes provided by the school educational administration system, the present course attendance rate and absence rate are calculated.

(2) Students' attendance. By analyzing facial morphological features in successive images, the number of students facing the blackboard, lowering their heads, and lying prone at their desks for long periods of time were identified. Then, the current class attendance rate, head down (looking at mobile phones) rate, and sleep rate were calculated.

(3) Other teaching information. Through the action recognition of continuous images, the characteristics of students “rushing to” the classroom door are judged, and then, the class time of the current course is obtained. Due to the diversity and complexity of students' movement behavior in the after-class, the judgment algorithm is not perfect, and the statistics is only an experimental function. Some statistics,

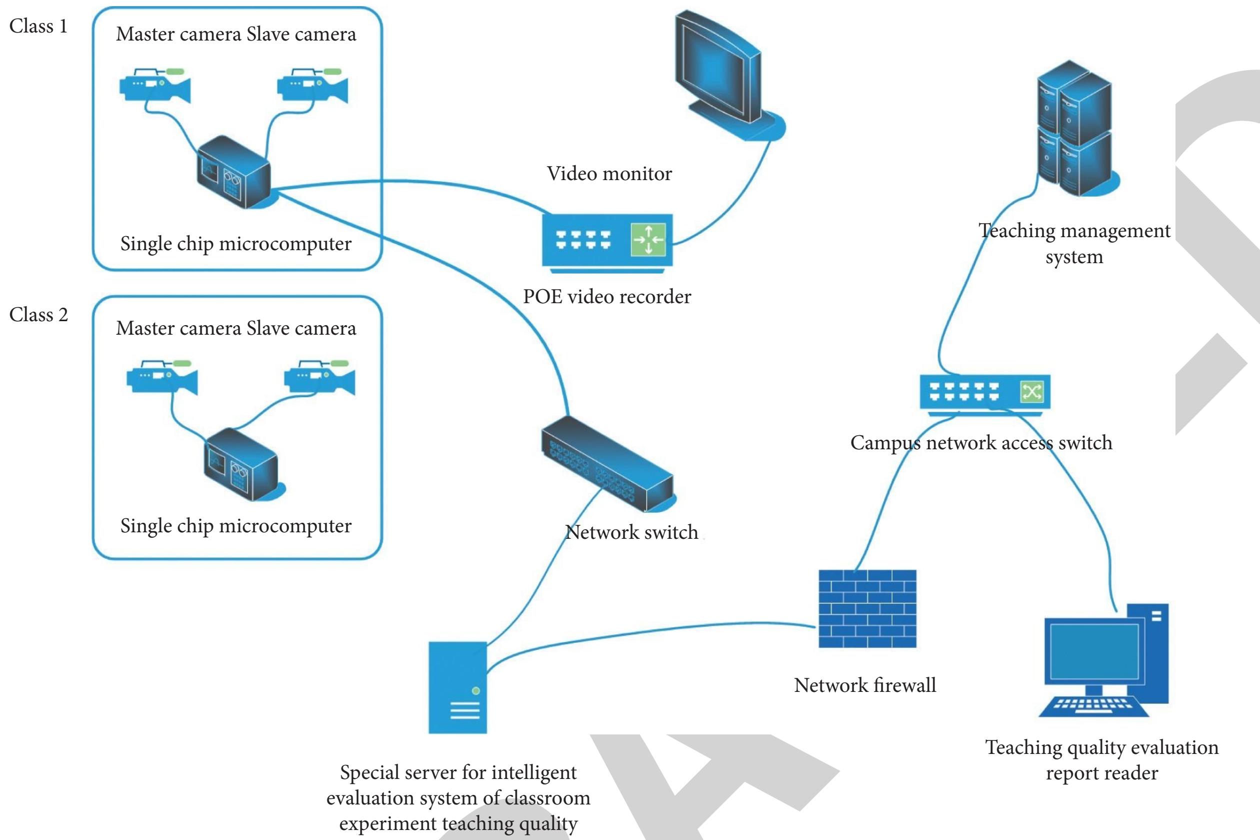


Figure 3: Software system block diagram of the proposed evaluation system.

such as absenteeism, tardiness, and mobile phone use, were part of the subsequent experiment.

The data of students' attendance rate, bowing rate, and abnormal attendance rate of each course in each classroom are counted and then sent to the special server of teaching evaluation system for further processing.

4. Experiment and Analysis

4.1. Validation of the Proposed Algorithm. In order to verify the effectiveness of the proposed algorithm, action recognition experiments are carried out on two large action recognition datasets, NTU 60 and NTU 120 [17]. NTU 60 and its extended version NTU 120 are currently the largest motion recognition dataset based on 3d human skeleton sequences. Each sample was an action sequence obtained by a Microsoft Kinect V2 camera in a restricted indoor environment. Each moment contains the 3-dimensional coordinates of 25 major human joints in the camera coordinate system. The NTU 60 dataset contains 56,880 samples from 60 action categories performed by 40 participants. The NTU 120 dataset extends the original sample by adding 57,600 samples. It expands the action categories to 120 and the number of participants to 106. Cross-participant recognition and cross-perspective recognition experiments were performed on both datasets. In the cross-participant recognition experiment, 50% participant samples were used as the training set, and the remaining 50% participant samples were used as the test set. In the cross-view recognition experiment, two of the samples were used as training sets, and the other sample was used as test sets. NTU 120 introduces more factors that affect perspective, including the height distance between the camera and the action participant, and extends cross-view recognition to cross-environment recognition (cross-setup). The two experiments investigated the learning ability of the algorithm model from different perspectives.

The validity of each part of the proposed algorithm was evaluated by testing on NTU 60 dataset and NTU 120 dataset. The performance of the proposed algorithm is compared. The confusion matrix of this algorithm is shown in Figures 4 and 5. It can be found by observation that the algorithm in this article is diagonally dominant on each class of NTU 60 and NTU 120 datasets. This shows that the algorithm in this article has achieved a good classification effect on these two datasets. But there are still behaviors that can be mislabeled because they are inherently so similar that even human perception can be hard to tell apart. This algorithm can extract the relation between objects well and reduce the error.

In order to compare the recognition accuracy of the proposed algorithm with that of other algorithms, further tests were carried out on the NTU 60 dataset and the NTU 120 dataset. Literature [6], Literature [7], and Literature [18] are selected as the comparison algorithms. The comparison results are shown in Tables 1 and 2. The best experimental data are indicated in bold. It can be seen that the method in this article achieves optimal results on both datasets. This further verifies the advantages of the algorithm in this article.

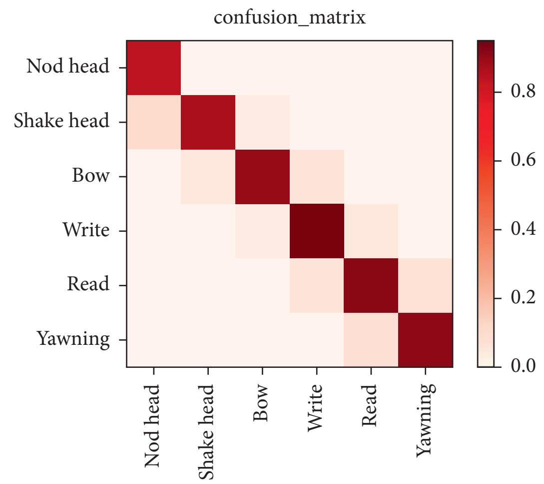


Figure 4: The confusion matrix of the NTU 60 dataset.

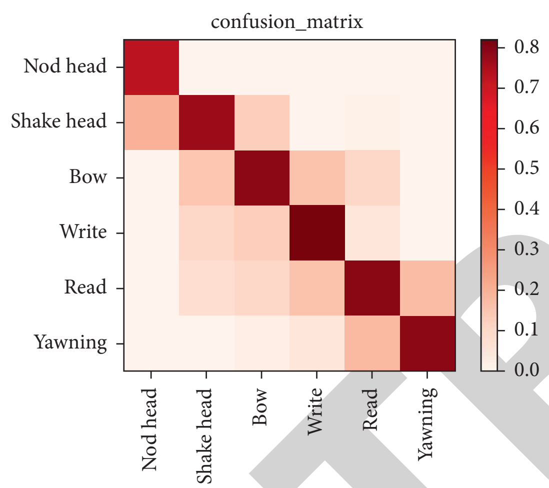


Figure 5: The confusion matrix of the NTU 120 dataset.

Table 1: Recognition accuracy of different methods on the NTU 60 dataset.

Methods	Cross-participant identification	Cross-view recognition
Literature [6]	82.2	88.5
Literature [7]	82.6	89.4
Literature [18]	81.7	91.6
Proposed	**84.5**	**92.4**

In order to compare the convergence performance of the algorithm, the convergence curve can be obtained by counting the loss function of the training process, as shown in Figure 6. It can be seen that Literature [6] is difficult to get convergence in some cases. The reason why the network is difficult to converge may be that its features are mainly calculated based on the low-order differential of the curve. It loses some sample information while maintaining invariance, so the distinction between samples becomes weak. However, this defect can be effectively compensated by combining invariant features with joint coordinates through

Table 2: Recognition accuracy of different methods on the NTU 120 data set.

Methods	Cross-participant identification	Cross-view recognition
Literature [6]	63.3	62.9
Literature [7]	72.8	75.4
Literature [18]	73.3	80.1
Proposed	**78.4**	**80.2**

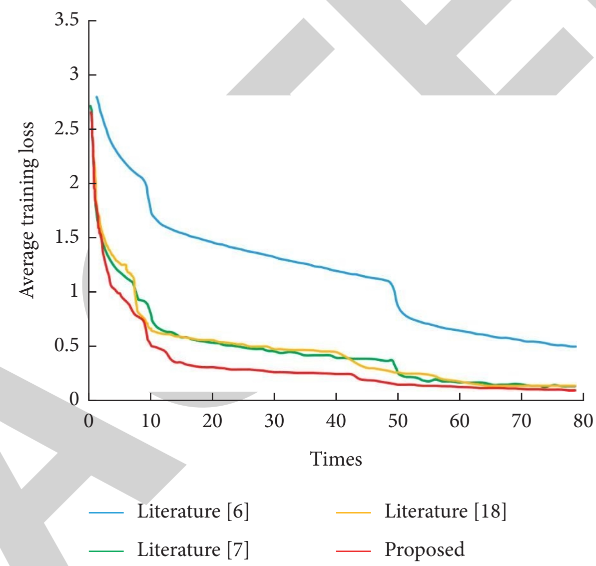


Figure 6: Convergence performance comparison.

channel enhancement. The algorithm in this article has faster convergence speed and more stable performance due to the fusion of spatiotemporal information.

4.2. Intelligent Evaluation of Foreign Language Classroom Teaching Quality. The special server of the foreign language course teaching evaluation system collects the data information sent by all cameras in the classroom. More intuitive statistics are processed by the back-end business logic module and stored in the database. The server system uses Windows Server + Tomcat + MySQL + Java software environment. The information stored in the database (data table) mainly includes the basic information of courses, classrooms, teachers, students, and colleges from the educational administration management system. At the same time, it also includes classroom situation, including course number, teacher, classroom number, and class time, which should be to the number of students, the number of students nod, the number of times of looking up, and the number of long time bow. In addition, there are separate data tables for high-frequency and specialized words told by teachers. The table also records the classroom, time, course name, and the list of high-frequency and professional words.

Educational administrators can call and view the classroom situation of each classroom in real time. In the video screen, the system marks the current student's head and difficult-to-identify areas with different color boxes. In addition, the server is responsible for summarizing the data information and generating various statistical reports for different users. The statistical reports include the evaluation report of teaching quality for teaching administrators, the evaluation report of curriculum teaching, the evaluation report of class style of study, and the evaluation report of teaching quality for teachers.

Through these reports, teachers' teaching and students' performance in class can be objectively reflected. It can help teachers understand the teaching situation after class and improve teaching methods. At the same time, it can also provide a fair and quantitative evaluation index for the teaching management and assessment of colleges and schools.

5. Conclusion

Foreign language courses play an important role in basic education. To analyze and judge students' behavior in a foreign language classroom, this article proposes a foreign language teaching quality evaluation system based on the fusion of spatiotemporal features. In order to describe interaction information effectively, a spatiotemporal modeling method was proposed, which combined intraframe interaction modeling design with interframe modeling. The potential relationship between joints is used for better identification to take full advantage of the spatial and temporal dependence of joints in the human body. The interactive skeleton graph is effectively represented, and then, the spectral convolution is used to extract spatial features. The algorithm in this article improves the accuracy of interactive action recognition, and experiments show the superiority of this method. The evaluation of classroom quality includes not only students' behavior in class, but also students' behavior outside class. Therefore, more factors will be considered in the evaluation system in the future to make the system more perfect.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interests.

Acknowledgments

This work was supported by the Dalian University of Foreign Languages.

References

[1] P. Ricciardi and C. Buratti, "Environmental quality of university classrooms: s," *Building and Environment*, vol. 127, pp. 23–36, 2018.

[2] S. A. D. Popenici and S. Kerr, "Exploring the impact of artificial intelligence on teaching and learning in higher education," *Research and Practice in Technology Enhanced Learning*, vol. 12, no. 1, pp. 1–13, 2017.

[3] A. M. Wu, K. Wang, J. S. Wang et al., "The addition of 3D printed models to enhance the teaching and learning of bone spatial anatomy and fractures for undergraduate students: a randomized controlled study," *Annals of Translational Medicine*, vol. 6, no. 20, 2018.

[4] A. Prati, C. Shan, and K. I. K. Wang, "Sensors, vision and networks: from video surveillance to activity recognition and health monitoring," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 1, pp. 5–22, 2019.

[5] J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, "Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition," *Pattern Recognition*, vol. 76, pp. 80–94, 2018.

[6] R. Cui, A. Zhu, G. Hua, and H. Yin, "Multisource learning for skeleton-based action recognition using deep LSTM and CNN," *Journal of Electronic Imaging*, vol. 27, no. 4, Article ID 043050, 2018.

[7] K. Cheng, Y. Zhang, C. Cao, L. Shi, J. Cheng, and H. Lu, "Decoupling gcn with dropgraph module for Skeleton-based action recognition," in *Proceedings of the Computer Vision–ECCV 2020: 16th European Conference*, pp. 536–553, Springer International Publishing, Glasgow, UK, August 2020.

[8] W. Li, X. Liu, Z. Liu, and F. Q. Du, "Skeleton-based action recognition using multi-scale and multi-stream improved graph convolutional network," *IEEE Access*, vol. 8, pp. 144529–144542, 2020.

[9] S. L. Happy and A. Routray, "Fuzzy histogram of optical flow orientations for micro-expression recognition," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 394–406, 2019.

[10] S. Wan, L. Qi, X. Xu, and C. Z. Tong, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Networks and Applications*, vol. 25, no. 2, pp. 743–755, 2020.

[11] H. Li, G. Zhu, L. Zhang, J. Song, and P. Shen, "Graph-Temporal LSTM Networks for Skeleton-based action Recognition," in *Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 480–491, Springer Cham, Nanjing, China, October 2020.

[12] C. Si, W. Chen, W. Wang, L Wang, and T. Tan, "An attention enhanced graph convolutional lstm network for skeleton-based action recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1227–1236, IEEE, Long Beach, CA, USA, June 2019.

[13] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, February 2017.

[14] H. H. Pham, H. Salmane, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin, "Spatio-temporal image representation of 3D skeletal movements for view-invariant action recognition with deep convolutional neural networks," *Sensors*, vol. 19, no. 8, p. 1932, 2019.

[15] Y. Li, T. Guo, X. Liu, and R. Xia, "Skeleton-based action recognition with lie group and deep neural networks," in *Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pp. 26–30, IEEE, Wuxi, China, October 2019.

[16] X. Gao, W. Hu, J. Tang, J. Liu, and Z. Guo, "Optimized skeleton-based action recognition via sparsified graph regression," in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 601–610, ACM, Nice, France, October 2019.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9865704, 1 page
https://doi.org/10.1155/2023/9865704

Retraction

Retracted: The Trend Prediction of the New Public Management Model Based on the Discrete Dynamic Evolution Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and name external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Zheng, "The Trend Prediction of the New Public Management Model Based on the Discrete Dynamic Evolution Model," *Security and Communication Networks*, vol. 2022, Article ID 3398392, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3398392, 14 pages
https://doi.org/10.1155/2022/3398392

Research Article

The Trend Prediction of the New Public Management Model Based on the Discrete Dynamic Evolution Model

Xiaoming Zheng

School of Culture and Law, Anhui Business College, Wuhu 241002, China

Correspondence should be addressed to Xiaoming Zheng; zxm@abc.edu.cn

Received 7 January 2022; Revised 14 February 2022; Accepted 23 February 2022; Published 22 April 2022

Academic Editor: Chin-Ling Chen

At present, Western theoretical circles have gradually reached a goal-oriented consensus on "civic participation in public management." This study provides an in-depth discussion of this synchronic view. At present, China's political and economic environment, government, and citizens' quality have undergone positive changes. The Chinese government has accelerated the practice of public management reform. Of course, there are also regional imbalances in the process of improving the public management mechanism in China. At the level of citizens, government, and society, there is a mismatch between supply and demand. With the increasing demands of citizens, the contradictions and crises in social life and public management are still intensifying. After clarifying these "risks" and "lag," the study makes recommendations for the practice of "citizen-participatory public management" in China. This paper studies from three dimensions of citizen's subjectivity, interpersonality, and sociality and specifically explores organizational flexibility, flattening, and diversity. The study expects to build a cooperative relationship between citizens and the environment and then build a "citizen participation" government. "Public management" has not been carried out in China for a long time, and its practicality is still limited. In this new management mode, realizing the benign interaction between citizens and the government is the key to the research. An important breakthrough point is to continue to nurture and develop community organizations (nonprofit organizations). This is because, by nurturing and developing community organizations, citizens can be provided with areas to realize their potential and achieve their aspirations. At the same time, this will also contribute to the multidimensional and diversified development of the organization. Most importantly, community organizations make public organizations more adaptable to the adjustment and integration of the external environment.

1. Introduction

As a new school of science, public management aims to help stakeholders acquire the knowledge, skills, and strategies needed to solve public problems, meet citizens' needs, and handle public affairs, to create a responsible, efficient, and fair government. For the handling of social and economic affairs, it is impossible to achieve Pareto optimality only by private supply and transactions. Providing and guaranteeing public services have become an important function of the government to make up for market deficiencies. While criticizing the narrowness of traditional public administration, the new public administration advocates responding to the democratization requirements of the society. The focus of public policy in the new era is on the implementation of policies, and the orderly advancement of policies is inseparable from the implementation and advancement of government administrative agencies. The introduction and management of private enterprises requires the effective protection of government public policies. The perspective is still limited to the government's system. It continues to adhere to the government's single subjectivity without changing the government's belatedness, financial, and efficiency crisis. . Therefore, the new public management that introduced the efficiency mechanism of private enterprises came into being [1–3]. While the new public management simplified the original organization process and improved internal efficiency and external response, it also made the citizenship singular, consumers and public managers [4–6]. Businesses of all kinds need to streamline their workflow. At present, the world is in the environment of globalization, accelerated information, and flattening of social structure. The predicament faced by enterprises has further intensified. There is a contradiction between the public's demand for public services and the government's supply. The needs of the public are diversified, but the supply provided by the government is relatively simple. There is a contradiction between the public and the government. This contradiction is the mismatch between everyone's demand for public services and the government's supply [7–9].

In the face of the above problems, public management cannot be resolved by looking back to the traditional bureaucracy or by softening management in morality. Enterprises need to pay more attention to the interaction with the external environment, so as to ensure that enterprises can achieve "fair and just" presentation results. Specifically, the implementation of public management policies should not only emphasize the rationality of its technology but also pay more attention to its democracy and openness to the outside world. Public management policies must focus on the component of public participation. Take public participation as the basic guarantee for building a good quality of life. There is a contractual relationship between the government and the private. The concept and practice of the boundary between subject and object is "citizen-participatory public management" [10–13].

In Europe and the United States, Western theoretical circles have gradually reached a consensus on citizen-participatory public management and have gradually become a trend. In daily practice, citizens interact with the government through open elections. At the same time, this trend has also brought a certain impact to the domestic people. Moreover, China is in the initial stage of a major transformation from government-mobilized public management to citizen-participatory public management. Regarding the reform methods of public management, there are disagreements in the theoretical circles. In the process of public management transformation and evolution, due to the "uniqueness" of its national conditions in the world, China has shown a fragmentary evolution process in the public management transformation, and an intermediate form has emerged rights protection response public management mode. At present, China is in the initial stage of cooperation between residents and the government and co-governance and management of grassroots communities. Regarding the discussion of public management policy reform, there are still some differences in the academic circles. Under the current stage of public management transformation and evolution, China's own national conditions are "unique" in the world, and the overall public management work still faces certain challenges. A new public management model is needed to effectively respond to existing problems. This middle-of-the-road model is reflected in different social classes responding differently. The public faces the separation of constitutional rights from concrete guarantees. People use different platforms to organize effectively. The government pushes the improvement of public administration through the reverse way. This behavior eventually evolved into an act of civic engagement. This behavior has also become a new model of public management [14–16].

From a practical point of view, although my country's socialist basic system provides a fundamental guarantee for citizens' participation in public management, there is a mismatch between supply and demand at the level of citizens, the government, and the society. Moreover, with the continuous rise of citizens' demands, it will aggravate the contradictions and crises in social life. If citizens' wishes and rights cannot be resolved through normal channels, they will often adopt unusual ways of participation [2, 17, 18]. In some parts of our country, such as environmental protection, demolition of prefectures and cities, land expropriation, and restructuring of state-owned enterprises, because the public cannot appeal for their own interests through normal channels, extreme irrational behaviors also occur from time to time. In the research on public management in Chinese academia, there are currently three characteristics. First, it is still in the stage of systematic introduction and sorting out of western new public management. The second is the concept of "governance" that has emerged in the West since the 1990s, especially the process from "good governance" to "good governance." There are many translations and introductions, but when it is applied to Chinese issues, it stays more in narration and introduction. However, there is a lack of effective discussion on the corresponding practical mode of public management. The third is that, in the study of China's public management model, western resources emphasized and the locality of knowledge ignored.

In this regard, the significance of this study lies in the following. First, it is proposed that citizen-participated public management is a way to implement or realize the concept of "good governance." In view of the current situation, the improvement of citizens' participation in public management is also an important means of building a harmonious society in our country. The continuous improvement of the national governance system also supports my country's modernization drive from another aspect. At present, China is facing both international and domestic pressures. The China's development model has stood the test of time. The China's public management model has been a huge success. The China's management model has withstood the dual tests of domestic historical factors and international demonstration effects. China is now generating a rights-defense response management model. This model is now on a scale, but it is not yet a sufficient condition for citizen participation in public management. This study analyzes how to move towards citizen-participatory public management based on the various positive and negative factors presented by rights-defense response public management. Regardless of whether it is an international or domestic background, citizen participation in the management of public affairs has become a trend of the times. Therefore, this study proposes that citizen-participatory public management has the significance of the times. In addition, citizen participation in public management is not only an important means of embodying people's democracy but also an effective way to improve government management, increase the efficiency of administrative democracy, and optimize the allocation of public resources [19–21]. Therefore, this study has a certain practical significance and practical value based on this topic. The research logical structure of this study is shown in Figure 1.

2. Existing Research Results and Theoretical Basis

2.1. Related Research Progress. Compared with the government, how society participates in public management is a new and continuous concept in academic research. In modern European and American countries, the category of "society" is equivalent to "civil society" and has rich civic characteristics. It is centered on the concepts of "citizenship" and "citizenship." The entire history of political development in the West can be seen as a controversy surrounding whether "citizenship" is recognized, how to exercise it, and the extent of its exercise. In the past 20 years, Western scholars have begun to reflect on the new public management [4, 22–24]. They believe that the management reform plan in the 1970s and 1980s was "dominated by efficiency and economic value, and efficiency can only be ranked third." The values "such as fairness, justice, representativeness, and participation are either removed from the agenda or seen as a stumbling block on the road to high productivity." As a result, democratic values such as fairness, justice, and citizen participation have been weakened, and they have fallen into a crisis of legitimacy in public administration. In addition, the introduction of a competition mechanism has neglected the cooperation and coordination between departments, which has brought about a fragmented institutional structure.

Therefore, they advocate a "post-new public management" reform, emphasizing that while retaining the value of "efficiency," public management pays more attention to the democratic constitutional implications of the government's responsiveness, representativeness, and citizen orientation. Although they put forward civic-oriented theoretical proposals, their awareness of future management models is still relatively vague. Western scholars have conducted a detailed analysis of how citizens participate in management. The background of its theories has a consensus on the interaction between citizens. The government needs to pay attention to the micro-level administrative process. The government should strengthen participation and management behavior in the administrative process and guide the public to take appropriate measures. There are differences in the policy process and the degree of relevance of citizens' participation in democratic values [5, 6, 14, 25]. For example, in the book "Citizen Participation in Public Decision-making: New Skills and New Strategies for Public Managers" written by American scholar John Clayton Thomas and translated by Sun Baying, it lists the criteria for judging the effectiveness of citizen participation in detail. Public managers choose what scope and degree of citizen participation in the process of public policy formulation and implementation to provide operability guidelines. It can be seen that the research on citizen participation in public management by Western scholars is not only limited to theoretical analysis, but has actually penetrated into reality.

In the late 1960s, as the phenomenon of citizen participation in urban planning became more and more common, many scholars who studied urban planning began to take citizen participation as their research object. After that, from the 1970s to the 1990s, many books on citizen participation published on topics such as environmental policies, antipoverty campaigns, community planning, and model city projects. In these works, scholars have increasingly connected citizen participation and poverty eradication, democratic decision-making, environmental protection, sustainable development, and other fields for research. Urban planning is no longer the only theme. At the same time, scholars' research on citizen participation is more comprehensive and in depth than before. In the areas of citizen participation, it involves issues such as land planning, forest resources, wetland resources, reservoir and water resources management, poverty eradication, community planning, and community development. In terms of citizen participation methods and techniques, it involves citizen hearings, citizen assemblies, advisory committees, citizen surveys, and citizen juries. Since the 1990s, there has been more and more research on citizen online participation. In terms of the value of citizen participation, it involves the impact of citizen participation, the relationship between citizens, government officials, and technical experts in the participation, the survey of citizen participation satisfaction, the relationship between citizen participation and democracy, and the representation of vulnerable groups and stakeholders, and citizens participate in the evaluation and other issues.

In recent years, with the development of information technology, especially geographic information systems, the modes of citizen participation have diversified. Some scholars have begun to explore how to use these technologies to promote citizen participation. Relatively speaking, the core content of this type of citizen participation research is relatively scattered, and a relatively complete theoretical system has not yet been formed. It has become a hotspot in the field of western citizen participation research in recent years [26–29]. Although Western academic circles have timeliness and forward-looking advantages in their historical and current research on public management, there is almost no relevant literature on the possibility and complexity of "citizen-participatory public management in China." Model analysis results are shown in Figure 2.

2.2. Corporate Government Model. The corporate governance model mainly involves the adjustment of the relationship between public administration and private administration. Western countries are accustomed to referring to the separation of legislative, executive, and judicial powers as public administration. Then, the activities of public administration departments to perform public functions in order to achieve public policy objectives are called public management. Correspondingly, the management activities of private enterprises and institutions are called private administration. There are many differences in purpose and method between public administration and private administration. Although this difference exists objectively, it is not insurmountable, especially the method and

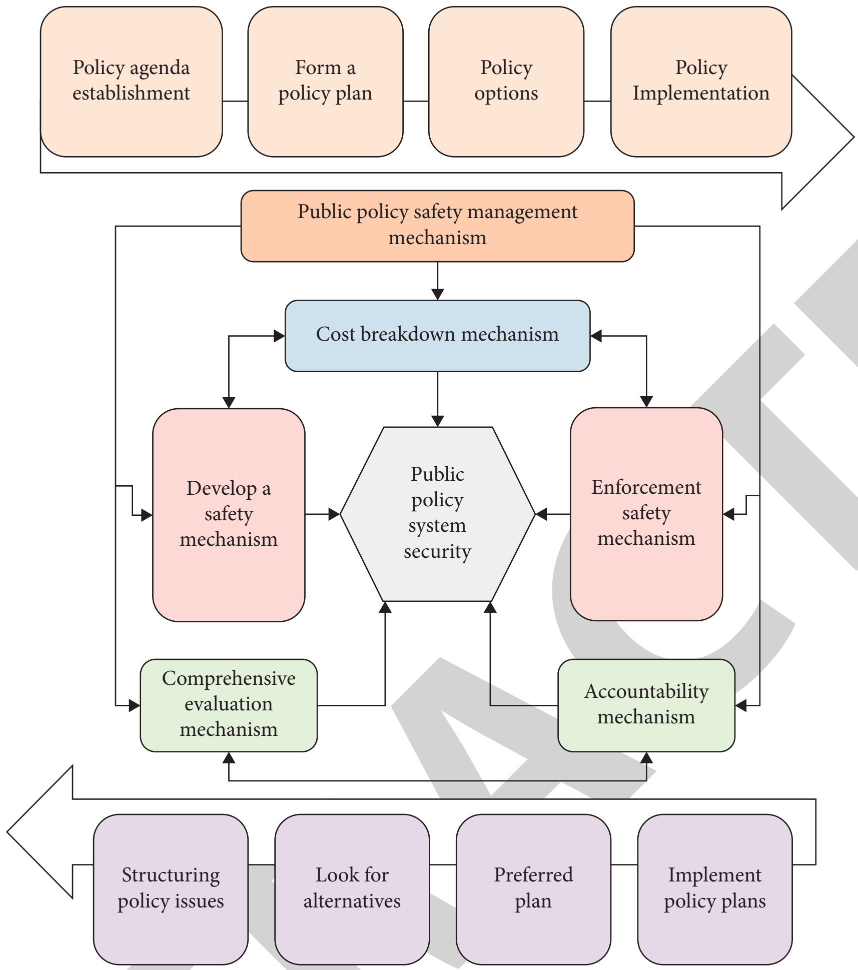


Figure 1: The research logical structure of this study.

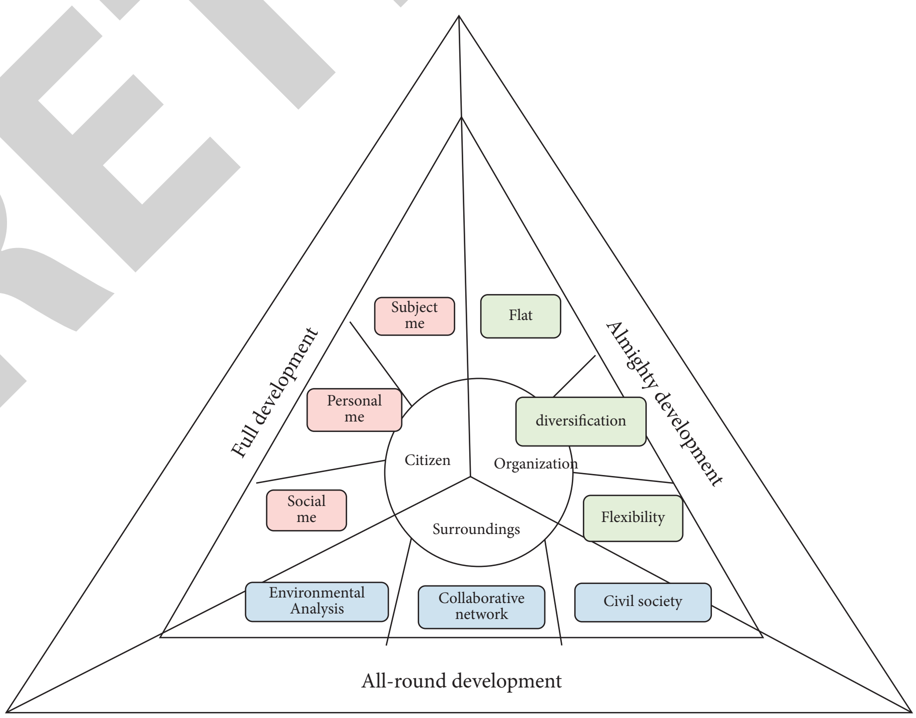


Figure 2: Citizen-participation public management model.

cross reference in the method. The issue is in a new round of government reforms. It has attracted widespread attention from scholars, among which the most influential are the views of American scholars David Osborne and Ted Gabler. Scholars believe that today's world needs to create a dynamic government, and corporate government is a major manifestation. Let us say we have a set of subjects $N = \{i = 1, 2, \ldots, n\}$. S is a subset of N, which represents the cooperative alliance that may be formed between the subjects and is the characteristic function of forming the cooperative alliance and represents the benefits obtained through the cooperative alliance:

$$\begin{cases} V(\phi_i) = 0, \\ V(S_1 \cup S_2) \geq V(S_1) + V(S_2), S_1 \cap S_2 = \phi_i, \end{cases} \quad (1)$$

$$\phi_i = \sum W\left(|S|\left[V(S) - V\left(\frac{S}{i}\right)\right]\right),$$

$$W(|S|) = \frac{(n - |S|)!\,(|S| - 1)!}{n!}, \quad (2)$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}\left|p_{ui} - q_{ui}\right|}{N}. \quad (3)$$

This study alleviates the impact of actual score differences between users on user trust relationships through a unified dimension [30]. The mean value of the scoring difference between the two users on the common scoring items is calculated as follows:

$$\varepsilon = \frac{\sum_{i \in I_a \cap I}\left(\left|c_{ai} - c_{bi}\right|\right)}{\left|I_a \cap I_b\right|}. \quad (4)$$

The constraint conditions are equations (2)–(4):

$$\begin{aligned} f(x_i, \omega) - y_i &\leq \xi_i + \varepsilon, \quad i = 1, 2, \ldots, l, \\ y_i - f(x_i, \omega) &\leq \xi_i + \varepsilon, \quad i = 1, 2, \ldots, l, \\ \xi_i, \xi_i^* &\geq 0, \quad i = 1, 2, \ldots, l. \end{aligned} \quad (5)$$

Coverage index calculates the ratio of predicted items to all unscored items, so as to measure the comprehensiveness of prediction. Assuming that h items are predicted, the calculation method of coverage is as follows:

$$\text{Cov} = \frac{h}{n}. \quad (6)$$

Recall index is also used to evaluate the system effect in the field of information retrieval. The larger Recall value is, the better the recommendation quality of the algorithm is.

$$\begin{aligned} \text{recall} &= \frac{\text{Hits}}{|\text{test}|} \\ &= \frac{|\text{test} \cap \text{Top} - N|}{|\text{test}|}. \end{aligned} \quad (7)$$

The corporate government has 10 characteristics. (1) The corporate government is a catalytic government. In the process of public management, the main task of the government is to formulate policies (to steer the helm) rather than to directly provide a certain service (planning). (2) The corporate government is a government owned by the community. The government should put its tentacles more deeply into the community. Through the interaction between the government and the citizens, the public can directly feel the care of the government and create citizens with a high degree of self-help ability. (3) The corporate government is a competitive government. Externally, the government promotes competition among private sectors for the provision of public services through competitive bidding. Internally, the government uses a work competition mechanism to promote quality competition among public departments with the goal of "winning customers." (4) The corporate government is a government with a sense of mission, and the government achieves this by removing outdated regulations and bad habits. The transformation from a traditional administrative culture that emphasizes procedures, processes, inputs, and rules to a new administrative culture that emphasizes goals, results, output, and missions and establishes a new responsibility mechanism and mission culture. (5) The government is a government that pays attention to results (effects). The government conducts performance management around its mission and implements new incentive mechanisms based on performance measurement. The interrelationships between model elements are shown in Figure 3.

2.3. Participatory Government Model. The participation model (Participation Model), also known as "Authorized Government Model," mainly involves the adjustment of the internal hierarchy of government organizations. In a government organization composed of high, middle, and grassroots levels, although the grassroots level is at the bottom of the organizational structure, it has an extremely important position. First of all, the large number of grassroots administrative staff is the most abundant part of the entire government organization with human resources. Secondly, the basic-level administrative agencies and personnel are the part of government organizations that have direct contact with the public. Their behaviors and activities directly affect the government's image, government prestige, and the public's evaluation of the government. Thirdly, due to the direct contact between basic-level administrative agencies and personnel and the public, they must make timely decisions on problems that arise; otherwise, they may intensify conflicts, which make basic-level administrative agencies and personnel turn from passive executives of administrative decision-making and become the creator of administrative decision-making. Facts have proved that a large number of decisions are not created by politicians or senior civilians, but by basic-level administrative agencies and personnel. However, in traditional government organizations, the basic-level administrative agencies and personnel are only regarded as the obeys of orders and the executors of decisions.

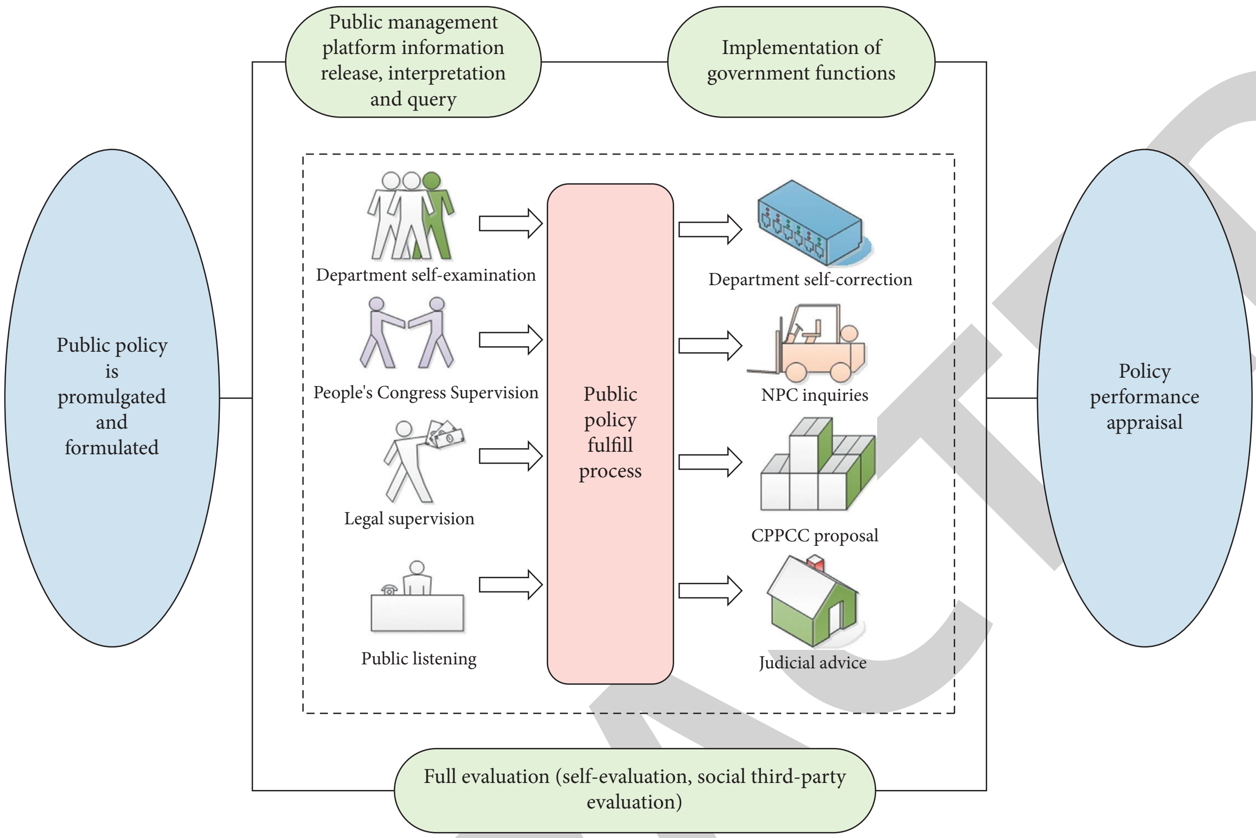


FIGURE 3: The flowchart of public policy promotion.

Among them, B_i is used as the scale element corresponding to the ith evaluation in the data set B. Through the dataset B, the data membership vector representing the injury of the athlete can be effectively integrated into a scalar. The formula is expressed as

$$V = r_j \times B. \tag{8}$$

Shape the general functional relationship between the output y of the injury model and the input $x_1, x_2, \ldots, x_n$. The Kolmogorov–Gabor polynomial is as follows:

$$\begin{aligned} y &= f(x_1, x_2) \\ &= a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2. \end{aligned} \tag{9}$$

And treat each of the monomials as m input models in the original structure of the modeling network:

$$\begin{aligned} v_1 &= a_0, \\ v_2 &= a_1 x_1, \\ v_3 &= a_2 x_2, \ldots, \\ v_6 &= a_5 x_1 x_2. \end{aligned} \tag{10}$$

As a generalization of the ordinary linear model, GLM introduces connection function in the model in order to fit some nonlinear relationships. The model can be expressed as

$$g(\xi) = g(\sigma) + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n, \tag{11}$$

where $g(\sigma)$ is the connection function, $\sigma = E(Y)$:

$$g(\xi) = a + f_1(X_1) + f_2(X_2) + \ldots + f_n(X_n). \tag{12}$$

Due to the passive decision-making power of the grassroots organizations, the entire government organization appears to be lacking vitality and slow to act, and it is difficult to obtain higher administrative efficiency, better administrative benefits, and affirmative social evaluation. In response to these situations and decentralization of power (especially decision-making power) to grassroots administrative organs and personnel, to develop grassroots human resources, mobilize grassroots staff's work enthusiasm, and reshape the purpose of the government.

2.4. Flexible Government Model. The flexible government model mainly involves the adjustment of the relationship between government organizations and government employees. Under the traditional administrative system, once a government employee enters a government organization, he can serve for life as long as he is willing and has no major work errors. The tenure of government employees is of great value for cultivating government employees' loyalty to the government, accumulating the work experience of government employees, improving the efficiency of government work, and maintaining the continuity of public policy implementation, but it nourishes the inertia of government employees. It stands still. In response to this situation, the flexible government model advocates increasing the flexibility of government organizations, introducing a "temporary employee system" in government organizations, setting up temporary institutions (such as some special committees and project teams), and hiring temporary personnel to complete administrative tasks. It is a new administrative task. Once the task is completed, these temporary institutions and personnel will be abolished. This approach can improve the state of government organizations and keep government organizations up to date with changes in the administrative environment. The ability to respond quickly can also avoid the negative impact of the "permanent industry" system on government employees and enable government employees to maintain keen judgment and active innovation and enterprising spirit.

2.5. Deregulation of the Government Model. The deregulating government model, also known as "deregulated government model" and "nonregulated government model," mainly involves the adjustment of the relationship between government and civil servants and government and the public. On the one hand, in traditional administrative organizations, civil servants are bound by strict organization and cumbersome regulations. Moreover, the work efficiency is low, which affects the overall interests of society. On the other hand, due to the cumbersome rules and regulations of the public sector, the complicated administrative procedures, and the excessively high cost of government public services, it not only caused the government's bureaucracy and corruption but also increased the burden on the public and the opposition and friction between the government and the public. For civil servants, the relaxation of government regulations by abolishing excessive and complicated administrative regulations and administrative procedures breaks through the traditional practice of bureaucracy. It treats decision-making as the privilege of politicians and gives government civil servants more contingency decision-making opportunities. The flexibility of implementing the rules and regulations allows it to maximize its creativity and work enthusiasm and thus benefits the overall interests of society. For the public, the relaxation of government regulations reduces administrative procedures, enables the public to obtain simple and fast services from the government, and better coordinates the interest relationship between the government and society.

3. The Chinese Government Management Model from the Perspective of New Public Management

The basic models or propositions of new public management in western countries have improved the level of public management in western countries to a considerable extent in practice, providing us with a new perspective for innovating our government management models in the new era. Some of these proposition and practice experiences are worth learning from.

3.1. Mode Change of Government Management Concept. The government management concept shows the government's behavior orientation to a certain extent and affects the government's management efficiency. To change the concept of the new government, it is necessary to establish the following three concepts.

Establish a market concept. Establishing a market concept means taking the market as the guide and introducing a competitive mechanism in government management. Current competition has expanded to competition between governments. The competitiveness of governments is mainly reflected in the quality of public products and the ability to provide services. In the supply of public services, the government should establish a market concept, introduce market mechanisms, give full play to the role of market mechanisms in resource allocation, reduce the cost of public service provision, and achieve market-oriented supply of public services by absorbing the participation of social forces. The diversification of public service subjects establish and improve the public service supply system. Through the introduction of market-oriented enterprise operation methods, the efficiency of public services and the quality of public services are continuously improved. Model analysis results are shown in Figure 4.

Establish a service concept. New public management advocates treating the public as "customers" and public service adheres to "customer orientation." Government management should change the government's official standard of focusing on power over responsibilities and management over service. The power-based concept is to establish a service concept that regards the public as "customers" and is "customer" oriented in the new era. In the process of conducting public management and fulfilling government functions, the government should be guided by the needs of citizens, adhere to the people-oriented principle, strive to improve the level of public services, and meet the growing public service needs of the broad masses of people. Establish a corresponding responsive mechanism to give citizens more choices and decision-making powers, and establish a "customer-driven" mechanism for citizens to evaluate and supervise the government's provision of public services. At the same time, government officials must also change their working methods and strive to improve their ability to serve the people, especially leading cadres at all levels "must firmly adhere to the view that leadership is service. Model analysis results are shown in Figure 5.

Establish a quality concept. New public management advocates the implementation of clear target control and performance evaluation and pays attention to the results and output of government management. Establishing the concept of quality is to require the government and its staff to change the concept of focusing only on process, investment, less results, and quality in the process of performing its government functions in the past. Control and measure the results of government management through the implementation of target management and performance evaluation, to achieve the best results with the least investment, and the final results must also focus on quality, and the results are connected with everyone involved in this management and

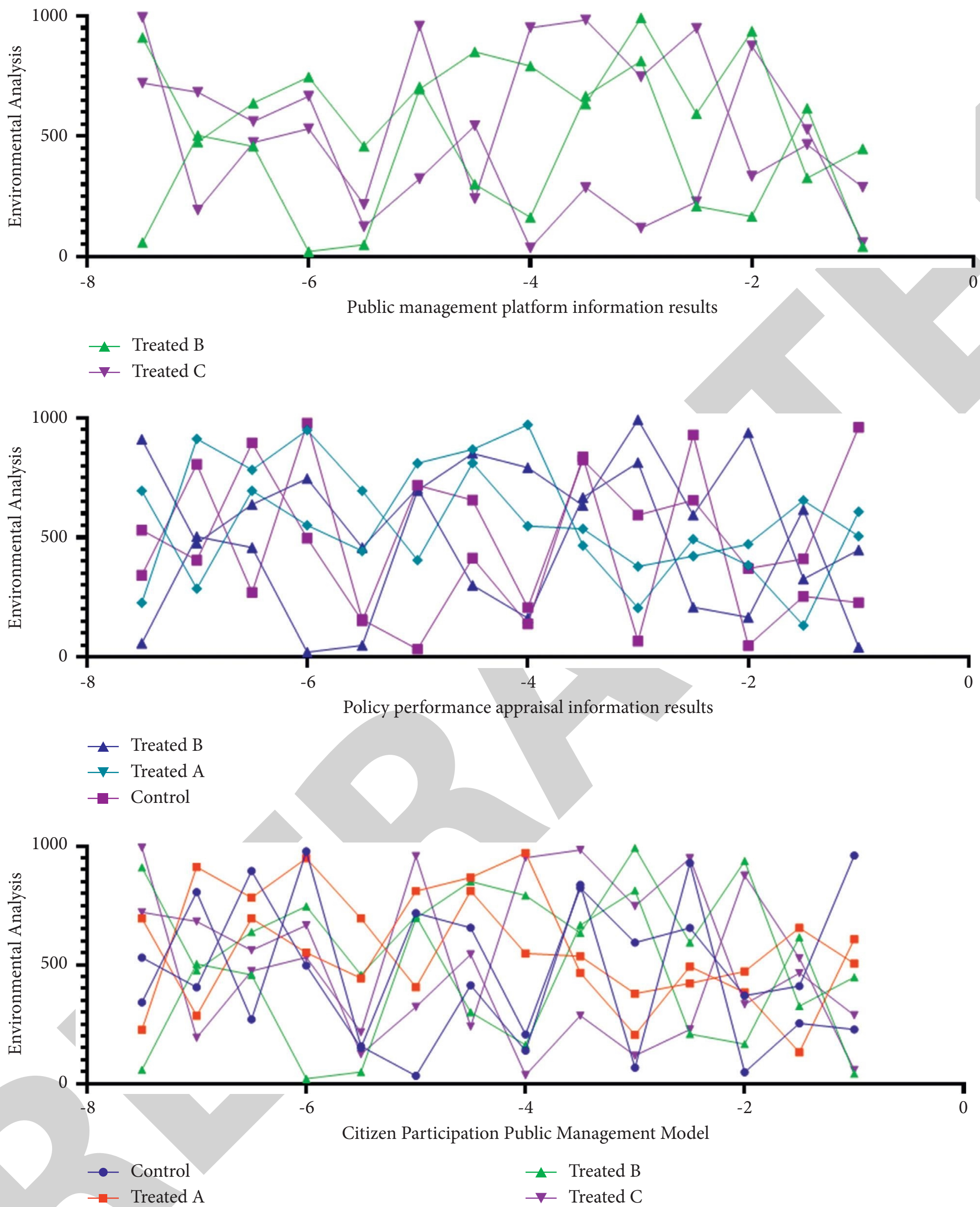


Figure 4: Analysis of the self-cognition model of subjective consciousness.

carry out performance appraisal. Model analysis results are shown in Figure 6.

3.2. Continue to Deepen the Reform of the Government Management System. Transform government functions. New public management has repositioned the government's functions, advocating that the government's management function should be steering instead of paddling, formulating policies rather than implementing policies, and implementing a market-oriented government function transformation. Through a market-based mechanism, the market's effective allocation of resources can be brought into full play, the government's intervention in the microspheres should be reduced, and the functional relationship between the government and the market, enterprises, and society should be straightened out. The phenomenon of "offside," "dislocation," and "absence" of government functions under the economic system. Establishing government functions for macroeconomic regulation, market supervision, social management, and public services under the socialist market economy are

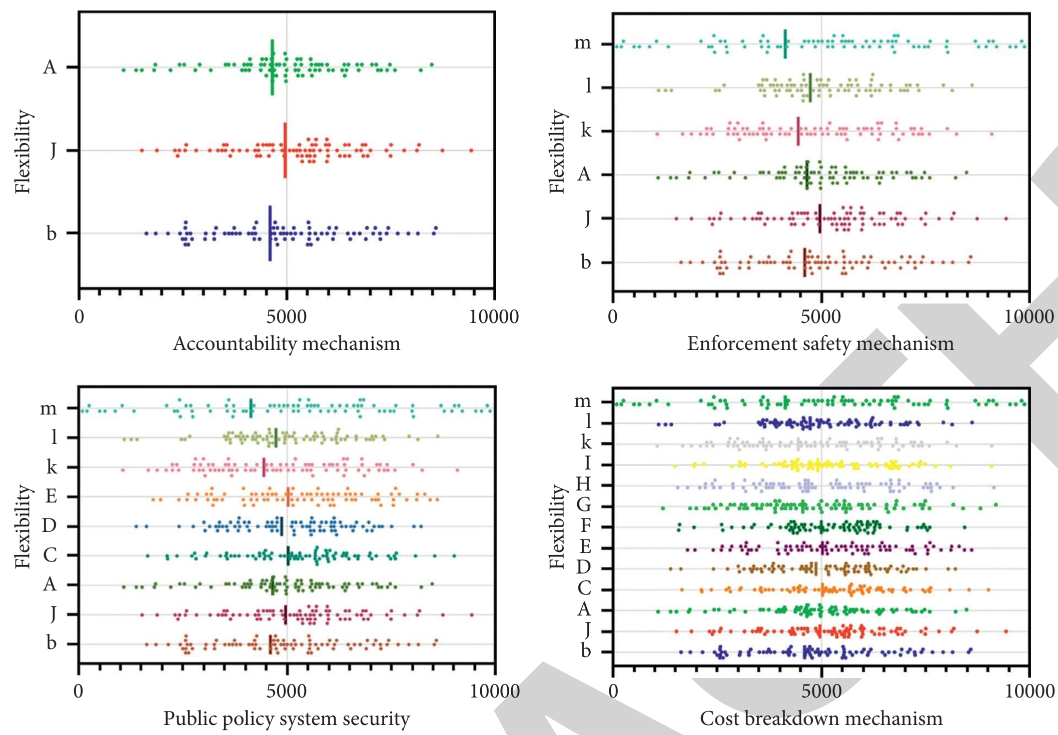


FIGURE 5: Analysis of the interpersonal relationship interaction model.

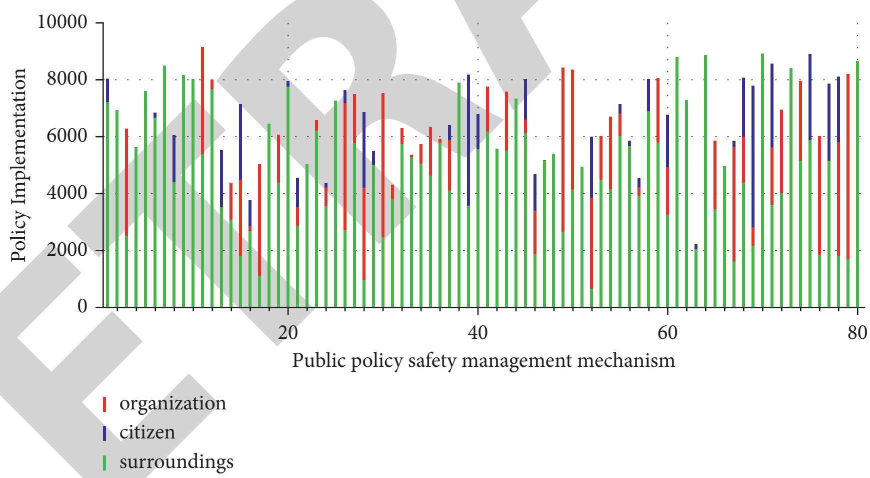


FIGURE 6: Establishment of social belief and model analysis.

obtained well treated. Model analysis results are shown in Figure 7.

The government needs to push forward with the reform of internal organizational structures. Enterprises need to improve the organizational structure system. According to actual needs, enterprises need to learn from the modern organization theory in the field of new public management. Enterprises need to reform and reengineer their own organizational structure. Enterprises need to establish a flat organizational structure that adapts to the current social and political system. In this way, the management level of the government reduced and the width of government management expanded. Through the application of modern information technology, the smooth transmission of information within the government and between the government and the outside world can be realized and the slowness of information feedback can be reduced. At the same time, actively promote the division of power and responsibility between the central and local governments and establish a relatively stable and flexible dynamic equilibrium relationship between the central government and local governments through decentralization and authorization. Model analysis results are shown in Figure 8.

Improve the level of public services. The monopoly of the supply of public services in my country has led to the low quality and efficiency of public services. The new public management introduced the competition mechanism into the government's public service field, breaking the

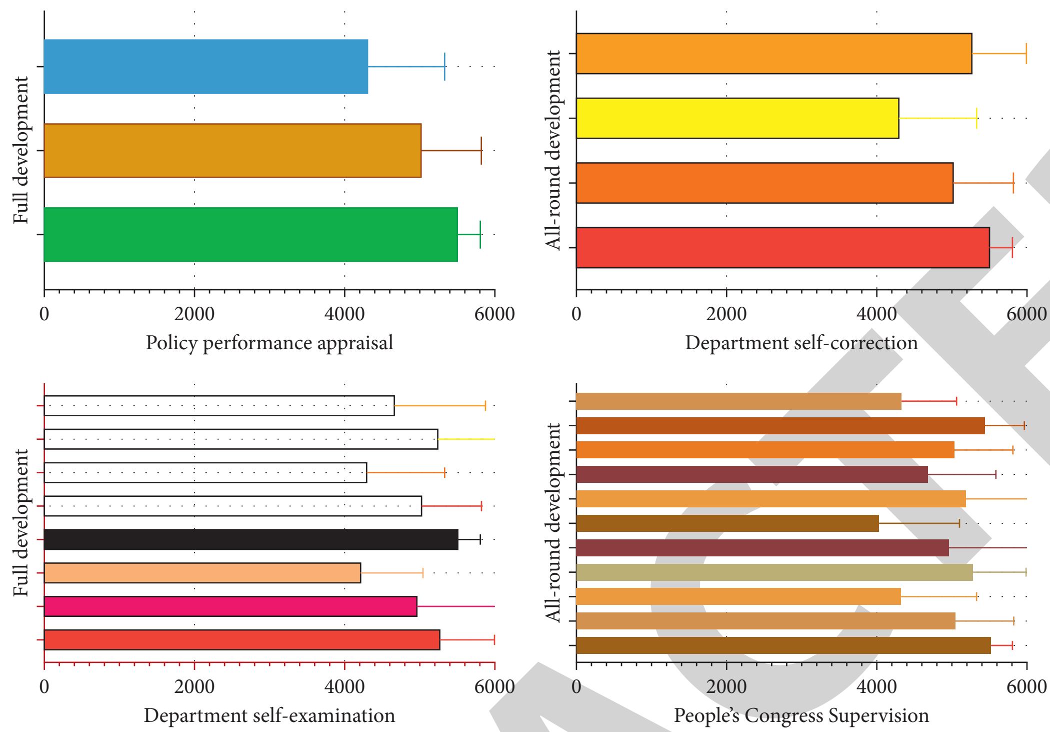


Figure 7: System supply model to ensure citizen participation in public management.

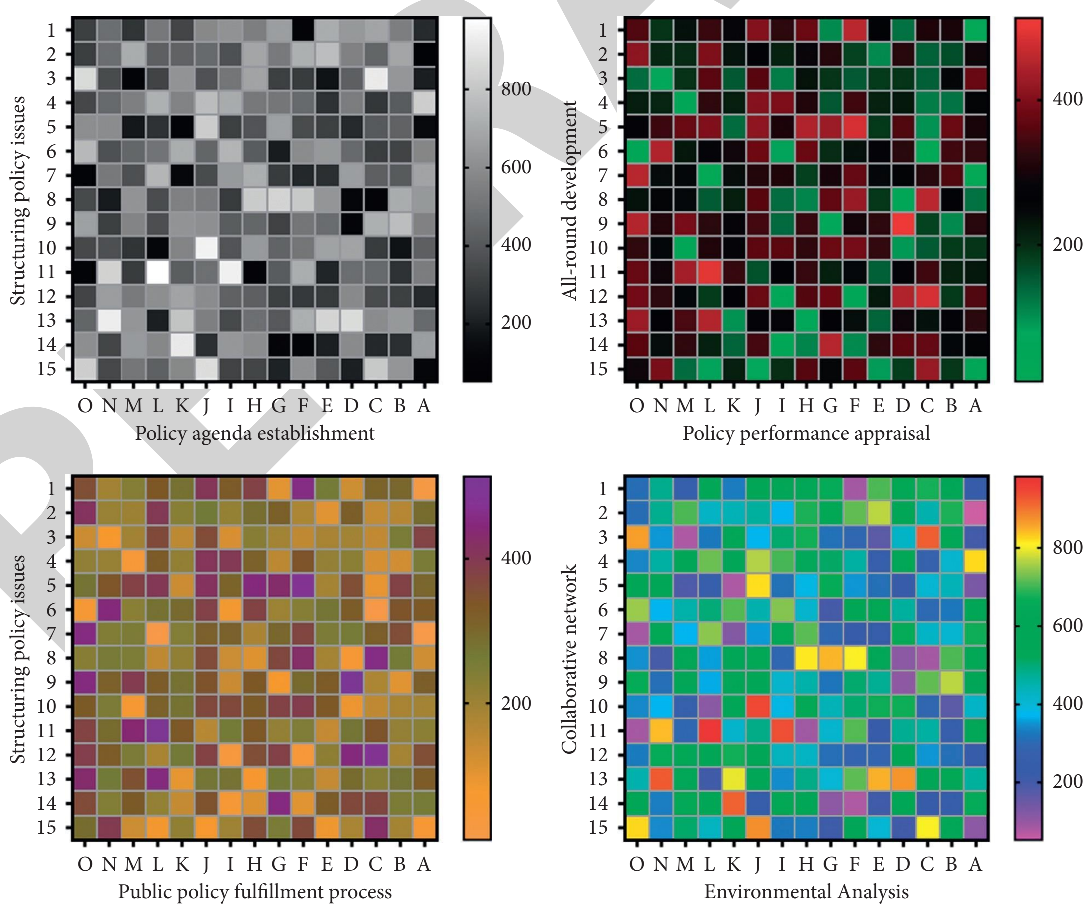


Figure 8: Improve the organization model of citizen participation public management.

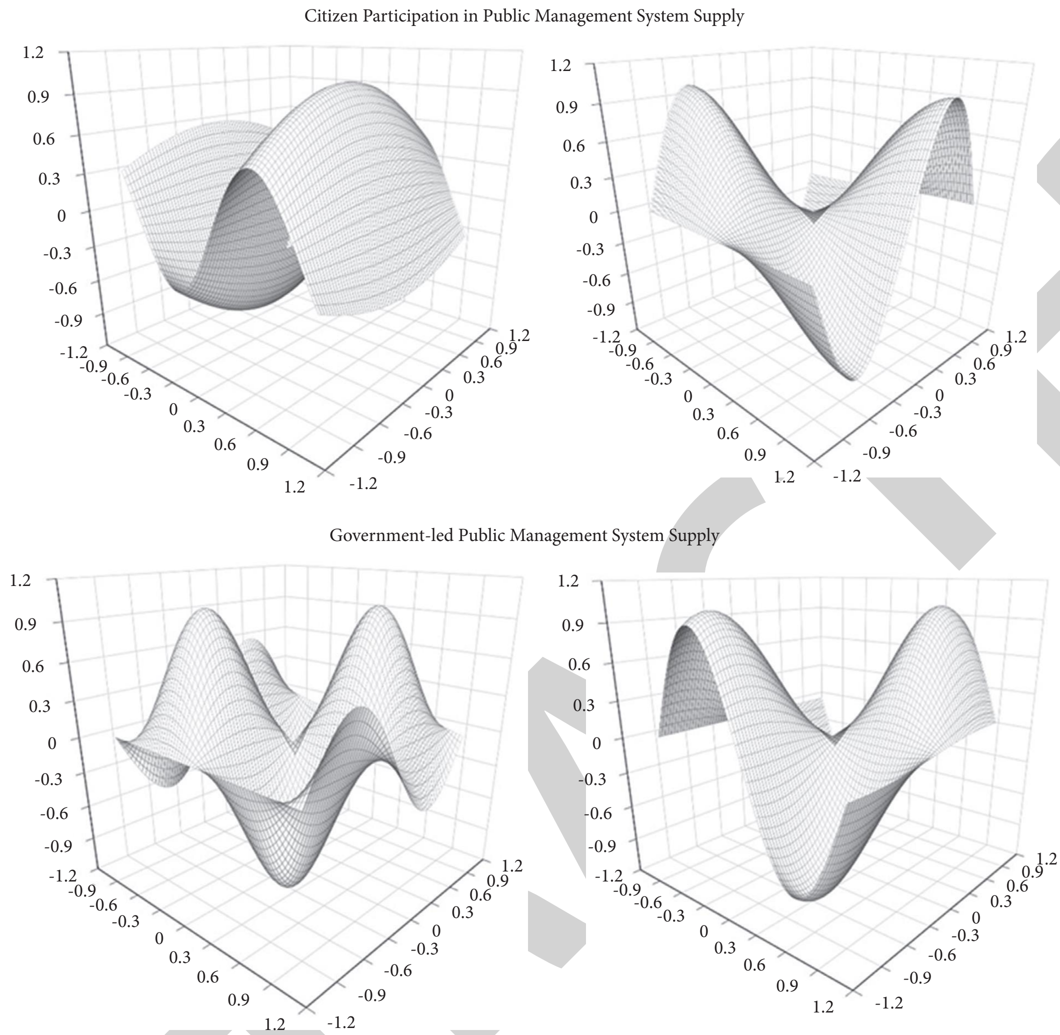


Figure 9: Analysis of the role of mass organizations on citizen participation in public management.

government's exclusive monopoly of providing public services. Through the competition mechanism, establish a diversified public service supply body, give full play to the role of the market mechanism in resource allocation, reduce the cost of public goods supply, absorb the participation of social forces and funds, and reduce the burden on the government. Improve the efficiency and quality of public service supply. Model analysis results are shown in Figure 9.

3.3. Innovative Methods of Government Management. Actively improve the methods and techniques of government management. We can introduce more mature methods and technologies such as target management, performance evaluation, and cost accounting, to government departments in business management. Establish management methods and technologies that adapt to the actual government, such as the establishment of performance appraisal mechanisms, target appraisal mechanisms, and administrative cost control mechanisms for government departments. New public management advocates introducing some management methods and technologies of enterprises into government departments. However, not all corporate management methods are completely suitable for government management, but the scientific nature of corporate management, emphasis on market demand, and customer feedback can be used for reference by public management so as to effectively improve the efficiency of government management. Model analysis results are shown in Figure 10.

Promote the construction of electronic government. New public management advocates the use of information technology to serve the organization and management of the government and the establishment of an electronic government. Promoting the government's own reform and construction through modern information technology and improving the government's management and service capabilities have become an important means of improving government management methods. Use information technology to establish and improve administrative management information systems including support information systems and management information systems. Establish more convenient information transmission and communication channels through network communication channels and methods, and provide technical support for the flattened

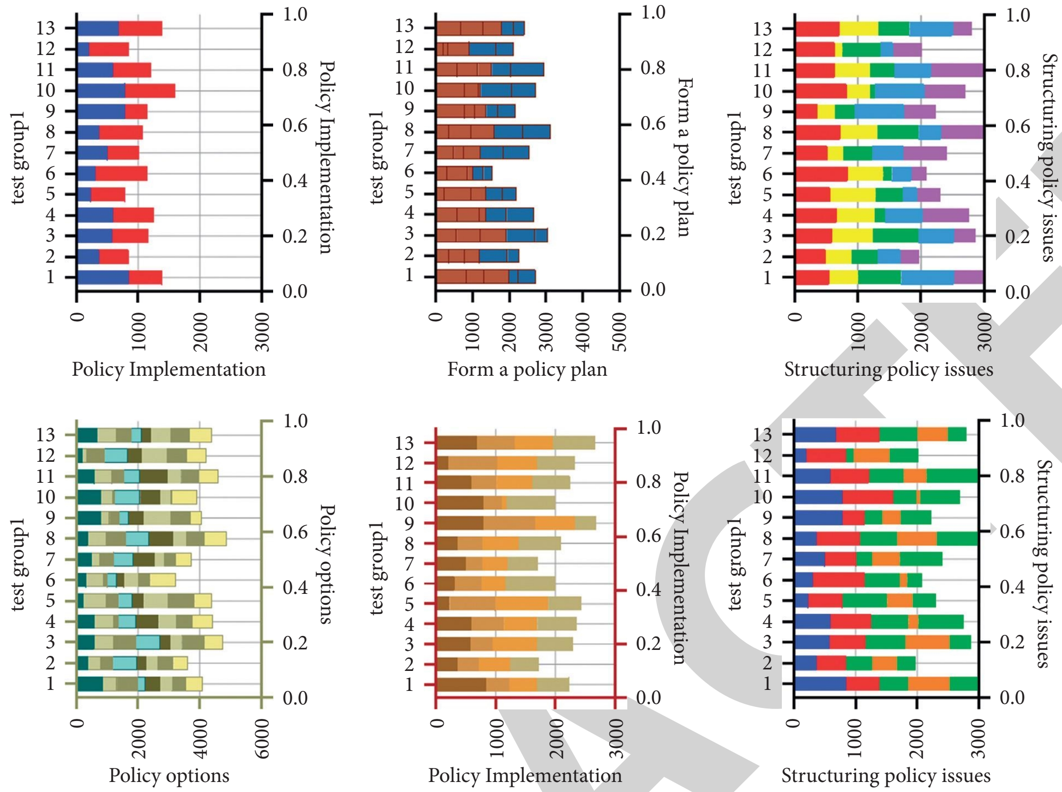


FIGURE 10: Analysis on the interaction mode of citizens, mass organizations, and government.

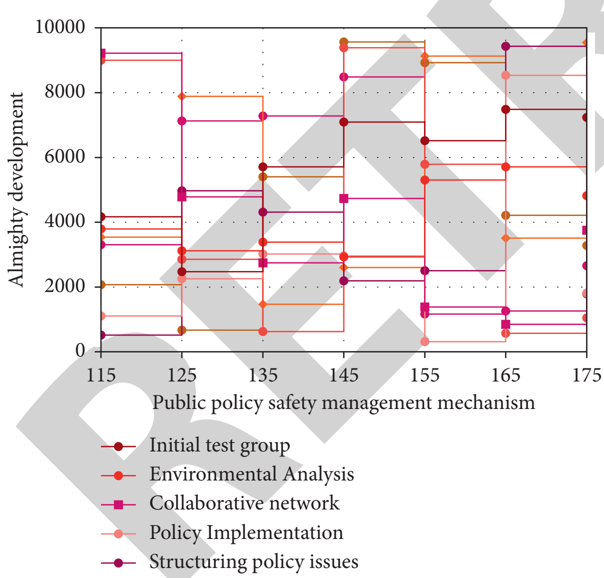


FIGURE 11: Public policy safety management mechanism.

government organization structure. By advancing the construction of e-government affairs, we will realize the digitization of government affairs processing, enhance the transparency of the government and public, and protect the people's right to know. Model analysis results are shown in Figure 11.

Western New Public Management has improved the administrative efficiency of Western countries to a certain extent, strengthened the competitiveness of Western countries in the international society, and provided a new perspective for us to innovate government management models. In the process of innovating the government management model, we should learn from some valuable insights and practices of Western New Public Management and, at the same time, pay attention to combining the actual situation of the current Chinese government, seeking truth from facts, and actively and steadily promoting the innovation of government management mode.

4. Conclusion

Organizations can not only draw resources from the environment but must also adapt to the complexity and uncertainty of the environment. The functioning of social organizations can be said to be positive feedback on environmental inputs and outputs. Therefore, members of the organization should view the development of society from a dynamic and developmental perspective. On the one hand, individuals should have a clear position on themselves, including their own strengths and weaknesses. On the other hand, pay attention to changes in the external environment and establish your own systemic values. The comprehensive development of civic participatory public management involves a wide range, not only an organization but also from the internal organization to many areas of society. Organizational patterns in this society have formed many new network-like structures. In the continuously developing social network structure of social organizations, it is necessary to maximize the practice of public management of citizen participation. Through vigorously cultivating and developing social organizations, we will realize the comprehensive and healthy development of society.

Social organizations can not only grasp real-time information but also understand the dynamic changes of the environment. Therefore, in the process of social governance involving citizen participation, the government should give full play to the role of community organizations. Community organizations can investigate the living habits of residents in detail and propose effective management methods. In addition to collecting relevant information, they also need to maintain good interaction with relevant groups and jointly establish a cooperative network for cooperating to implement the system. With the increasing demands of citizens, the contradictions and crises in social life and public management are still intensifying. After clarifying these risks, the study puts forward suggestions for the practice of “citizen participation in public management” in our country. This study starts from the three dimensions of citizen’s subjectivity, interpersonally, and sociality. The study explores the flexibility, flatness, and diversity of social organizations to build synergies with the environment. The realization path of “public management” in China is relatively complicated. Under this new management model, the society specifically realizes the benign interaction between citizens and the government. The breakthrough point of this cooperative model is the continued cultivation and development of community-based organizations (nonprofit organizations). This is because, by cultivating and developing community organizations, not only can it provide citizens with a space to realize their potential and realize their ideals but also contribute to the multidimensional and diversified development of the organization. Most importantly, community organizations make public organizations more adaptable to the adjustment and integration of the external environment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this study.

Acknowledgments

This work was supported by the domestic study visit and training program for outstanding young backbone teachers in Colleges and universities of Anhui Provincial Department of Education (gxgnfx2021055), construction project of high-level “double qualified” teacher training base in innovation and development pilot zone of Anhui Provincial Department of Education (WJ-PTZT-171), and “building a high-level double division team” project under the double high plan (2020sgxm04).

References

[1] A. Edrees, H. Abdelhamed, S. W. Nho et al., “Construction and evaluation of type III secretion system mutants of the catfish pathogenEdwardsiella piscicida,” *Journal of Fish Diseases*, vol. 41, no. 5, pp. 805–816, 2018.

[2] K. Alexiou and J. Wiggins, “Measuring individual legitimacy perceptions: scale development and validation,” *Strategic Organization*, vol. 17, no. 4, pp. 470–496, 2019.

[3] Z. Yang and L. S. C. Pun-Cheng, “Vehicle detection in intelligent transportation systems and its applications under varying environments: a review,” *Image and Vision Computing*, vol. 69, pp. 143–154, 2018.

[4] X. Li, Y. Wang, Q.-H. Wang, Y. Liu, and X. Zhou, “Modified integral imaging reconstruction and encryption using an improved SR reconstruction algorithm,” *Optics and Lasers in Engineering*, vol. 112, no. 6, pp. 162–169, 2019.

[5] J. M. Cairney, K. Rajan, D. Haley et al., “Mining information from atom probe data,” *Ultramicroscopy*, vol. 159, no. 1, pp. 324–337, 2020.

[6] Z. Wang, H. Ren, Q. Shen, W. Sui, and X. Zhang, “Seismic performance evaluation of a steel tubular bridge pier in a five-span continuous girder bridge system,” *Structures*, vol. 31, no. 1, pp. 909–920, 2021.

[7] S. Schnelle, J. Wang, R. Jagacinski, and H.-J. Su, “A feedforward and feedback integrated lateral and longitudinal driver model for personalized advanced driver assistance systems,” *Mechatronics*, vol. 50, pp. 177–188, 2018.

[8] J. Barrena-Martinez, M. López-Fernández, and P. M. Romero-Fernández, “The link between socially responsible human resource management and intellectual capital,” *Corporate Social Responsibility and Environmental Management*, vol. 26, no. 1, pp. 71–81, 2019.

[9] H. Aguinis, Y. H. Ji, and H. Joo, “Gender productivity gap among star performers in stem and other scientific fields,” *Journal of Applied Psychology*, vol. 103, no. 12, pp. 1283–1306, 2018.

[10] E. M. A. Ahmed, “A hydrologic-economic-agronomic model with regard to salinity for an over-exploited coastal aquifer,” *Journal of Geosciences*, vol. 12, no. 12, pp. 1–12, 2019.

[11] G. Han and W. Fu, “The lateral tracking control for the intelligent vehicle based on adaptive PID neural network,” *Sensors (Basel, Switzerland)*, vol. 17, no. 6, pp. 25–33, 2017.

[12] L. Ye and T. Yamamoto, “Modeling connected and autonomous vehicles in heterogeneous traffic flow,” *Physica A: Statistical Mechanics and Its Applications*, vol. 490, no. 40, pp. 78–81, 2018.

[13] H. B. Gao and X. Y. Zhang, “Longitudinal control for mengshi autonomous vehicle via cloud model,” *IOP Conference Series: Materials Science and Engineering*, vol. 320, no. 1, pp. 324–340, 2018.

[14] S. Kumar Dwivedi, R. Amin, V. Satyanarayana, and C. Rashmi, “Blockchain-based secured event-information sharing protocol in internet of vehicles for smart cities,” *Computers and Electrical Engineering*, vol. 86, no. 1, pp. 1–9, 2020.

[15] J. E. Varajao, “A new process for success management-bringing order to a typically ad-hoc area,” *Mod. Proj. Manag.*vol. 5, no. 3, pp. 94–99, 2018.

[16] C. M. Kang, S.-H. Lee, and C. C. Chung, “Multirate lane-keeping system with kinematic vehicle model,” *IEEE*

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9791489, 1 page
https://doi.org/10.1155/2022/9791489

Retraction

Retracted: Design and Application of BP Neural Network Optimization Method Based on SIWSPSO Algorithm

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 17 November 2022

Security and Communication Networks has retracted the article titled "Design and Application of BP Neural Network Optimization Method Based on SIWSPSO Algorithm" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] L. Chu, "Design and Application of BP Neural Network Optimization Method Based on SIWSPSO Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 2960992, 9 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2960992, 9 pages
https://doi.org/10.1155/2022/2960992

Research Article

Design and Application of BP Neural Network Optimization Method Based on SIWSPSO Algorithm

Lina Chu

Department of Basic Education, Chongqing Creation Vocational College, Chongqing 402160, China

Correspondence should be addressed to Lina Chu; rena_chu@163.com

Received 25 February 2022; Revised 19 March 2022; Accepted 31 March 2022; Published 21 April 2022

Academic Editor: Chin-Ling Chen

BP neural network method can deal with nonlinear and uncertain problems well and is widely used in the construction of classification, clustering, prediction, and other models. However, BP neural network method has some limitations in fitting nonlinear functions, such as slow convergence speed and easy local optimal convergence rather than global optimal convergence. In order to solve the insufficiency, the optimization approach applying BP neural networks is discussed. This paper proposes a simplified PSO algorithm based on stochastic inertia weight (SIWSPSO) algorithm to optimize BP neural network. In order to test the effect and applicability of the method, this paper established a quality safety risk warning based on SIWSPSO-BP network and selected the detection data of intelligent door lock products for risk warning experiment. The experimental results show that the convergence speed of SIWSPSO-BP model was increased by two times and the accuracy of product quality risk warning reached 85%, which significantly improves the accuracy and learning efficiency of risk warning.

1. Introduction

ANN is a computational model which simulates biological neural network. ANN approaches the target function through repeated training, which has the characteristics of parallel processing, discreteness, and self-learning. This algorithm can carry out large-scale nonlinear operations on big data and is suitable for building learning and prediction models [1–3]. At present, the more mature neural networks include convolutional neural network and BP neural network. Due to the complexity of convolutional neural network, the advantages of BP neural network have been highlighted. It is widely used in many fields such as artificial intelligence, signal processing, and automation. It is the most widely studied and applied artificial neural network at present.

In recent years, many scholars have carried out extensive research on the application of neural network. The optimization of BP network [4] is also in progress. The main optimization is to find the algorithm of optimal network error, optimize network input and network parameters, etc. The methods of the global optimal solution to find the optimal error of the network include PSO algorithm [5], GA algorithm [6], and compression mapping genetics [7]. The input of network optimization is in the aspect of data feature extraction [8, 9], such as principal component analysis (PCA) method [10], 13-point feature extraction algorithm [11], and four-angle feature extraction method [12]. The optimization of network parameters includes the addition of error momentum term [13], the use of adaptive learning rate [14], and the use of adjustable activation function [15]. Literature [16] studied the BP neural network model based on the end echo reflection method, taking the probe K value and the sound path difference of the reflected wave corresponding to the highly cracked end as the input vector for network training. In literature [17], the improved genetic algorithm was used to conduct global search for the optimal weights and thresholds of the network, and BP algorithm was used for local optimization to obtain the predicted wind speed. Literature [18] introduced the double-layer evolution mechanism of cultural algorithm into particle swarm optimization algorithm and updated its own speed and position by learning the optimal particle. In literature [19], the adaptive chaotic particle swarm optimization algorithm is used to optimize the objective function. When the function solution falls into the local optimum, the chaotic search is used to guide the particle to search again.

BP network has some problems, such as slow convergence speed, sensitivity to initial value, local minimum, hidden layer number, and neuron number. A simplified particle swarm optimization (SIWSPSO) algorithm based on random inertia weight is proposed to explore the optimization method of BP neural network application. The backpropagation neural network is optimized. An experimental model was established to verify the method.

This paper mainly has the following innovations:

(1) The inertia weight is described by random variables, and the learning factor adopts asynchronous change strategy to replace the individual extreme value of each particle with the average value of the individual extreme value of all particles.

(2) The simplified particle swarm optimization algorithm based on random inertia weight is used to optimize and solve the global optimal value.

This paper consists of five main sections: Section 1 gives an introduction, Section 2 discusses related research, Section 3 presents the model design, Section 4 provides case application and analysis, and Section 5 concludes the paper.

2. Related Work

2.1. BP Neural Network. BP neural network consists of input node layer, hidden node layer, and output node layer. When information is propagated forward, input data is passed in from the input layer [20]. If the actual output of the output layer in the neural network does not conform to the expected output, the direction of reverse propagation of the steering error is reversed, and the weight of each node layer is correctly modified according to the error [21]. The operating principle of BP neural network is that the information is propagating forward and the error is propagating backward. The error signal is used as the basis for correcting the weights of each unit. Its network topology is shown in Figure 1.

M_{yx} represents the weights of the input node layer and the hidden node layer. M_{xl} represents the weights of the implicit node layer and the output node layer. p_x^N represents the input of the x-th neuron in the hidden node layer. q_x^N represents the output of excitation from x-th neuron. The current sample set is I, and $I_z = [I_{z1}, I_{z2}, \ldots I_{zt}]$ for any sample.

The actual output is Z_t, and the mathematical expectation D_z is the number of iterations of T.

(1) Network inputs sample I_z, which can be obtained from forward information transmission:

$$\begin{aligned} p_x^N &= \sum_{w-1}^{W} M_{yx} I_{zw}, \\ q_x^N &= f\left(p_x^N\right) = f\left(\sum_{w-1}^{W} M_{yx} I_{zw}\right), \\ p_y^Y &= \sum_{x-1}^{X} M_{xl} Q_x^X, \\ q_x^Y &= \oint\left(p_x^N\right) = \oint\left(\sum_{x-1}^{X} M_{yl} I_x^X\right), \\ Z_t &= q_x^Y == \oint\left(\sum_{x-1}^{X} M_{yl} I_x^X\right), \end{aligned} \tag{1}$$

outputting node-layer neural network error signal:

$$e_z(t) = d_z(t) - j_z(t). \tag{2}$$

The sum of errors at the output node layer is as follows:

$$E(t) = \frac{1}{2}\sum_{l-1}^{L} e_z^2. \tag{3}$$

The information is always passed forward from the first input node layer and ended at the output node layer, and the learning error *E(t)* is calculated.

(2) Error reverse transmission process is carried out when errors are generated. Error information is reversely transmitted from the output node layer backward to the input node layer. The weight and threshold of each layer are modified. Now, the partial derivative of the error to the weight in the algorithm is directly proportional to the correction of the weight in the output node layer and the hidden node layer [22]; that is,

$$\triangle M_{xl}(t) \propto \frac{\partial E(t)}{\partial M_{xl}(t)}. \tag{4}$$

Among them,

$$\frac{\partial E(t)}{\partial M_{xl}(t)} = \frac{\partial E(t)}{\partial p_{lL}(t)} \times \frac{\partial p_{lL}}{\partial M_{xl}(t)} = \frac{\partial E(t)}{\partial e_z(t)} \times \frac{\partial e_z(t)}{\partial Z_z(t)} \times \frac{\partial Z_z(t)}{\partial P_{yY}(t)} \times \frac{\partial P_{yY}(t)}{\partial M_{xl}(t)}. \tag{5}$$

Then, after the next iteration, the weight is adjusted to

$$M_{xy}(t+1) = M_{xl}(t) + \Delta M_{xl}(t). \tag{6}$$

At this time, the network error feedback propagation is completed, and the weight between each node layer is worth an iterative update. BP neural network algorithm needs

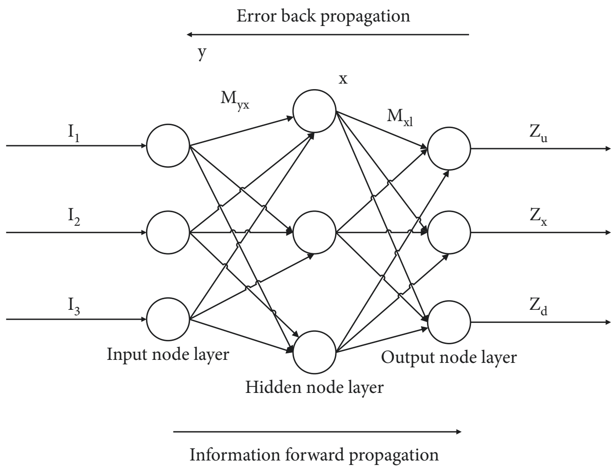


Figure 1: BP neural network structure model.

repeated iterations to make the error generated by learning converge to the expected accuracy [23].

2.2. Basic Principle of Gradient Descent Algorithm. Gradient descent method is also known as the fastest descent method, which searches for the weight and threshold of the optimal state along the direction of negative gradient. Now, the gradient descent algorithm is designed, and its strong local search ability makes up for the local search dilemma of BP neural network [24]. The operation is as follows.

For a sample of t features, it is expressed as a function.

$$b(i_1, i_2, \ldots, i_t) = \theta_0 + \theta_1 i_1 + \theta_2 i_2 + \cdots + \theta_t i_t. \tag{7}$$

θ_x ($x = 0, 1, 2, \ldots, t$) represents the coefficient on the model, and i_x ($x = 0, 1, 2, \ldots, t$) represents t variable parameters for each sample. Its loss function can be expressed as

$$y(\theta) = \frac{1}{2w} \sum_{l=0}^{w} \left(b\left(i_0^y, i_1^y, \ldots, i_w^y\right) - z^y \right)^2. \tag{8}$$

The smaller the loss function is, the better the fitting degree is.

The negative gradient of the loss function is calculated by

$$-\frac{\partial}{\partial \theta\ x} y(\theta) = -\frac{1}{w} \sum_{l=0}^{w} \left(b\left(i_0^y, i_1^y, \ldots, i_w^y\right) - z^y \right) i_x^y. \tag{9}$$

Multiply the step size α by the negative gradient of the loss function to get the current correction, which is $-\alpha\partial/\partial\theta\ xy(\theta)$, and update the current correction, which is $\theta_{\text{new}} = \theta_x - \alpha\partial/\partial\theta\ xy(\theta)$.

At this point, an iteration of gradient descent is completed, and the above process is repeated until the loss function tends to the minimum and the gradient descent iteration ends.

2.3. PSO Algorithm. Inspired by the foraging behavior of birds, particle swarm optimization (PSO) is a random optimization algorithm based on swarm intelligence. The main idea is to solve the problem through the learning behavior of individual particles and the cooperative interaction between groups. Due to its advantages of simple structure, strong parallelism, and fast convergence, PSO algorithm is mostly applied to solve multiobjective optimization, neural network training, image processing, and other problems [25]. However, PSO is also prone to premature convergence and local optima. In view of these problems, many scholars have put forward improvement methods from various aspects, that is, parameter adjustment, elite selection, algorithm mixing, and other strategies. Literature [26] had proposed a linear decreasing rule of inertia weight. On this basis, literature [27] improved the convergence and convergence speed of the algorithm by dynamically adjusting the strategy. In addition, some scholars made similar improvements on the acceleration factor of PSO and proposed the time-varying self-organizing PSO algorithm (HPSO-TVAC) [28]. In order to further improve the particle optimization ability, literature [29] applied the neighborhood optimal solution to the velocity updating formula and determined the learning intensity through the iterative process. In recent years, good results have been obtained by using topological structure changes to select learning models. As shown in literature [30], a new two-layer population structure was designed by using the double-difference mutation strategy, and elite particles were generated by inference mutation operation with two different control parameters to ensure population diversity. A particle swarm optimization algorithm with heterogeneous clustering (APSO-C) was proposed in literature [31]. K-means clustering was used to dynamically cluster the population, information was exchanged among clusters through ring structure, and learning samples were selected. Many studies have shown that both dynamic multipopulation structure [32] and adaptive learning framework can effectively improve the global search ability of the algorithm. It can overcome th

e precocious phenomenon of PSO to some extent. In terms of hybrid algorithm, some scholars also use genetic operator and simulated annealing operation to deepen local optimization, and combine differential evolution algorithms to modify the global optimal particle.

2.3.1. Basic Ideas of Particle Swarm Optimization. PSO algorithm originated from the study of foraging behavior of birds. There are some similarities between swarm foraging and optimization problem solving. Therefore, people simulate the biological principle of bird swarm foraging to make optimization decisions and find the optimal solution of the problem. The implementation process of the standard PSO algorithm is as follows.

Suppose t particles form a population in the W-dimension (that is, there are W function independent variables) search domain and t represents the population size. If the population is too small, the diversity of particle population cannot be guaranteed and the algorithm performance is poor. Although too large population can increase the efficiency of optimization and prevent premature convergence, it will undoubtedly increase the amount of calculation, resulting in too long convergence time, which is manifested as slow convergence. $I_x=(i_{x1}, i_{x2},...,i_{xW})$ ($x=1, 2,..., t$) is the position vector of particle I, and the particle dimension depends on the number of variables of the function to be optimized. Here, $i_{xz}\in[L,P]$ represents the value of particle X on the Z-th independent variable. In practical application, the value of each dimension of I is guaranteed to be within a certain range. This is equivalent to the domain of independent variables in function optimization problems. L represents the lower limit of the Z-th independent variable, and U represents the upper limit of the k-th independent variable. $Q_x=(q_{x1},q_{x2},...,q_{xW})$ is the velocity vector of particle X, and they are all W dimensions. $Q_{xz}\in[q_{\min}, q_{\max}]$, where $q_{\min}$ represents the minimum velocity of the particle in the z-dimension direction and $q_{\max}$ represents the maximum velocity of the particle in the z-th dimension. In each generation of optimization, the particle will adjust its flight direction and orientation according to the historical optimal position found by itself and the colony.

Remember that $U_x=(u_{x1}, u_{x2},...,u_{xW})$ is the position with the best adaptive value found by particle X itself. Remember that $U_a=(u_{a1}, u_{a2},...,u_{aW})$ is the optimal position searched by the whole particle swarm.

Let $f(i)$ be the fitness function, and the individual optimal position of particle I is as follows:

$$u_x(z)=\begin{cases} u_x(z-1), f(u_x(z))\geq f(u_x(z-1)),\\ i_x(z), f(u_x(z))<f(u_x(z-1)). \end{cases} \tag{10}$$

The global optimal position found by all particles in the group is as follows:

$$u_a\in\{[u_1(z),u_2(z),\ldots,u_W(z)]|f(u_a)=min(f(u_1(z),u_2(z),\ldots,u_W(z))\}. \tag{11}$$

When the x-th particle of the n-th generation evolves to the $(n+1)$-th generation, the velocity and position of the y-th dimension are calculated by the following evolution equation:

$$\begin{aligned} q_{xy}(n+1)&=m\times q_{xy}(n)+c_1\times r_1\times\left[u_{xy}(n)-i_{xy}(n)\right]+c_2\times r_2\times\left[u_{ay}(n)-i_{xy}(n)\right]\\ i_{xy}(n+1)&=i_{xy}(n)+q_{xy}(n+1). \end{aligned} \tag{12}$$

where m is the inertial weight, C_1 and C_2 are acceleration factors, and r_1 and r_2 are random quantities.

The velocity and position of each one-dimensional particle will be constrained within a range. In order to prevent the particle from escaping out of the solution space, if the boundary condition is exceeded, the following methods are adopted:

when $q_{xy}>q_{\max}, i_{xy}>i_{\max}, q_{xy}=q_{\max}, i_{xy}=i_{\max}$,

or

when $q_{xy}<-q_{\max}, i_{xy}<-i_{\max}, q_{xy}=-q_{\max}, i_{xy}=-i_{\max}$.

2.3.2. Influence of Parameters on Particle Swarm Optimization. The search performance of the algorithm is highly dependent on parameters. The algorithm involves three parameters: inertia weight m and acceleration factors c_1 and c_2. If the three parameters are set to constant values or linear changes, this will adversely affect the optimization and efficiency of the algorithm. Improper setting of the three parameters may cause the particle swarm optimization algorithm to evolve into a local optimization algorithm, or the particle swarm will lose its diversity at an early stage, resulting in premature convergence of the algorithm. In addition, in the early stage of optimization, it can improve the ability of particles to search the global optimal solution and make particles have higher velocity. However, when approaching the optimal solution in the later stage, in order not to make the particle velocity too high and deviate from the optimal location region, the particle misses the global optimal solution and falls into the local optimal solution. Therefore, when approaching the global optimal region in the later stage, the position updating amplitude should not be too large, and the particle velocity should be effectively adjusted and constrained. We should not ignore the fact that the particle may move out of the global optimal region due to excessive velocity in the later stage, which may lead to immature convergence of the algorithm. In view of the above reasons, the nonlinear shrinkage factor $\rho(n)$ is introduced into the PSO calculation.

In the process of algorithm search, the change of inertia weight value should meet the following requirements: the speed of early reduction is relatively slow, the inertia weight value is large, and the reduction is small, which is conducive to global exploration. The later stage is small, and the reduction speed is fast, which is conducive to the particle carrying out fine local search and effectively avoiding falling into local optimum. In addition, the changes of the two acceleration factors of the algorithm should meet the requirements that c_1 first becomes large and then small, and c_2 first becomes small and then large, so that the algorithm can give consideration to both local and global search.

3. Model Building

All models in this paper were trained and tested on a computer with Core I5-7500 CPU @3.40 GHz and 64 GB memory. The computer system is 64-bit Windows 10

Professional. All models are implemented with Matlab 2020a Deep Learning Toolbox framework.

3.1. SIWSPSO Optimized BP Neural Network. In BP neural networks, random initialization weights and thresholds usually make the network fall into local extremum points, which further affects its nonlinear fitting ability and network operation efficiency. Initial thresholds and weights are equivalent to particles in particle swarm optimization. The test error of BP neural network is taken as the fitness function. The time-domain peak value, center frequency, 3 dB bandwidth in frequency domain, upper limit cut-off frequency f_H, and lower limit cut-off frequency f_L of transmitted wave signal are selected as the input of neural network, and the expected value is the output of neural network. In this paper, the random inertia weight value is introduced to improve the PSO algorithm, and the application model of SIWSPSO-BP neural network is established. See Figure 2.

3.2. Product Quality Risk Warning Model. Product quality risk early warning model adopts qualitative and quantitative methods to conduct comprehensive risk analysis of products, and the evaluation steps are as follows:

(1) The unqualified test items are taken as risk factors, and the unqualified rate is divided into five levels of 20% each. According to Table 1, determine the x value.

(2) The impact of risk factors is divided into 5 levels and assigned to each level, as shown in Table 2.

(3) A number of experts in the field of product quality and safety and risk assessment are invited to assess the impact of items on risk occurrence according to expert analysis and evaluation, and the scores are given according to the instructions in Table 3. The average score is used as the rating basis to determine the j value.

(4) According to expert analysis and evaluation, determine the values of w and t ($w + t = 1$), divide the risk grade into different grades, and define the corresponding relationship between the risk grade and the quantitative value (K).

$$K = \sqrt{wi^2 + t j^2}. \tag{13}$$

In the formula, K is the risk grade value, w is the weight value of the possibility of risk factor, t is the weight value of the influence degree of risk, i is the possibility of the occurrence of risk factor, and j is the influence degree of risk factor.

(5) Calculate the risk level values according to steps 1–4, and build the risk level classification matrix according to the possibility of risk, as shown in Table 4.

(6) Select the risk factor with the highest risk grade value, and determine the risk grade of the risk assessment of the product according to the risk grade classification matrix.

3.3. Training Warning Model Based on SIWSPSO-BP. The product quality warning model was designed and trained based on SIWSPSO-BP algorithm, and input layer vector, hidden layer node number and weight, output layer vector and activation function were determined.

(1) Input layer vector. Product quality risk warning needs to input a variety of product quality testing items, which have different effects on product quality. Therefore, key testing items affecting product quality are extracted as input vectors.

(2) Hidden layer design. The hidden layer is designed as 1 layer with 4 nodes, which correspond to the risk evaluation of experts (the probability value of the occurrence of risk factor i, the influence degree of risk factor j, the weight value of the possibility value of risk factor W, and the weight value of the influence degree of risk factor t).

(3) Output layer vector. Conclusion of quality safety risk assessment: 3 represents high risk, 2 represents medium risk, 1 represents low risk, and 0 represents no risk.

(4) Activate the function. Select the unipolar S-type Sigmoid function [33]. The function is flat on both sides and continuously differentiable in the middle, being suitable for activation function.

4. Case Application and Analysis

Intelligent door lock is implemented through wireless network, NFC, modern biology or optics, remote control, and other technologies, in product security, user identification, and management of more intelligent and simple products. Smart door locks have security risks such as data leakage risk, replay attack, fingerprint cracking, and communication security. Based on the above SIWSPSO-BP neural network risk warning model, we carry out risk warning experiment for intelligent door locks.

We collected 210 product records of risk monitoring data of intelligent door locks from the risk monitoring database of Tianjin Quality Inspection Institute, with a total of 4320 test data items. 200 product records were selected as training samples and 20 as prediction samples. When conducting comprehensive rating, detection items such as antiviolent opening, RFID lockpick security, antidamage alarm, sensitive information protection, data encryption transmission, security scanning, decompilation, and network security are taken as input. The test result of the item is pass or fail. The detection results are shown in Table 3.

Initial setting of SIWSPSO-BP algorithm parameters was as follows: input: 21; hidden node number: 4; output node: 1; genetic population size: 20; genetic algebra: 40; crossover probability: 0.5; mutation probability: 0.3. The BP neural network model was established on Matlab, and the optimization model based on SIWSPSO-BP neural network was

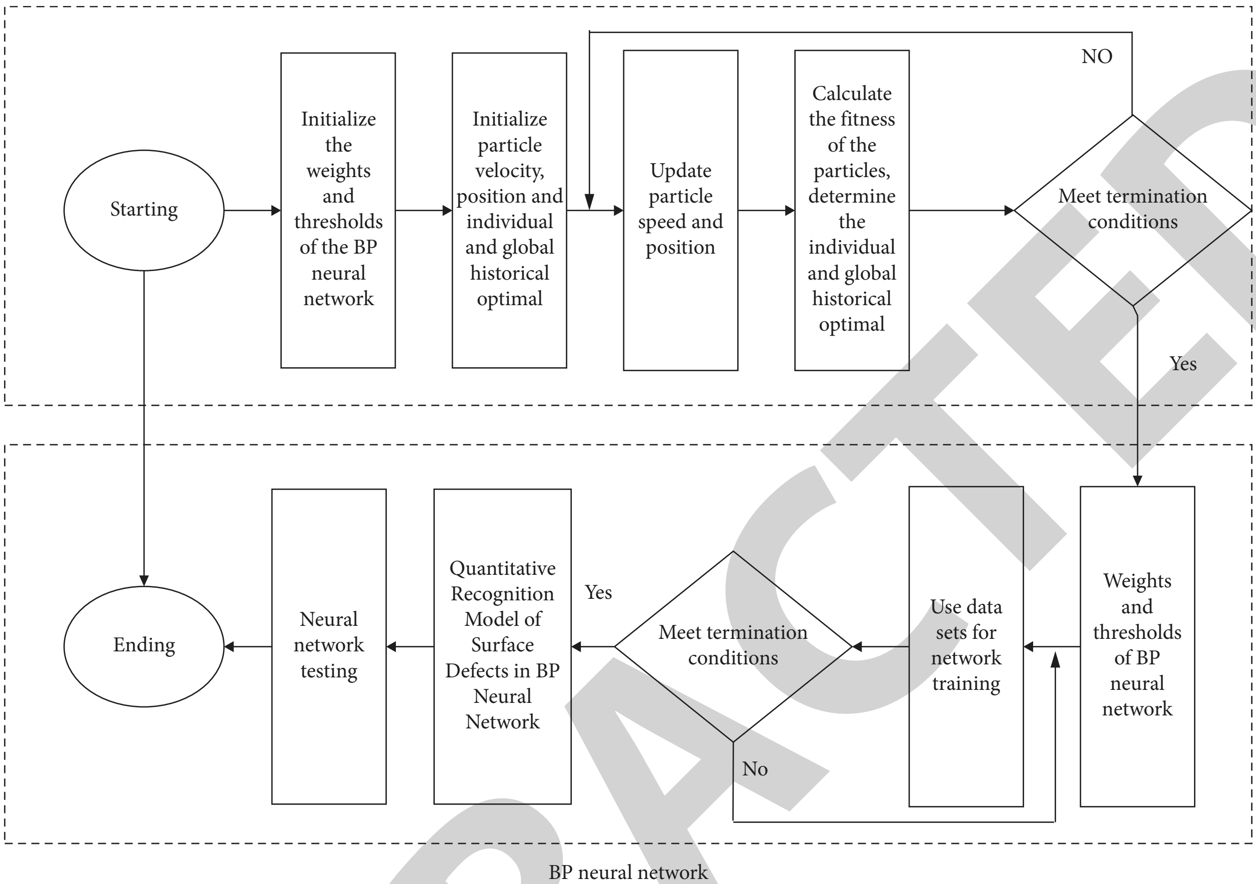


Figure 2: SIWSPSO-BP model.

Table 1: Classification of risk occurrence possibility.

Level	Feature description	Assignment
Very low	(0 ~ 20%)	(0 ~ 1)
Low	(20 ~ 40%)	(1 ~ 2)
Medium	(40 ~ 60%)	(2 ~ 3)
High	(60 ~ 80%)	(3 ~ 4)
Very high	(80% ~ 100%)	(4 ~ 5)

Table 2: Description of influence degree of risk factors.

Degree of influence	Description	Assignment
Key	Risk factor is the key factor leading to risk	(4 ~ 5)
Serious	The impact of risk factors is significant	(3 ~ 4)
General	The impact of risk factors is moderate	(2 ~ 3)
Small	Risk factors have little influence	(1 ~ 2)
Can be ignored	The impact of risk factors can be ignored	(0 ~ 1)

realized by programming. Training samples were used to train the model.

As shown in Figures 3 and 4, the BP neural network based on genetic algorithm has greatly improved the convergence times of training times, and the convergence speed of SIWSPSO-BP is twice that of BP, which improves the training speed. In terms of prediction error, the accuracy of SIWSPSO-BP in product quality risk warning reached 85%, compared with the accuracy of 75% of BP model before optimization, and SIWSPSO-BP neural network also has obvious advantages.

Table 3: Summary analysis table of networked intelligent door lock project detection results (section).

			Product information			Test item								
Number	Sample number	Name of manufacturer	Sample name	Specifications	Trade mark	Anti-violence opening	RFID lock safety	Anti-damage alarm	Password entry test	Sensitive information protection	Data encryption transmission	Security scanning	Decompiling	Network security
1	W01814501800	Company 1	Intelligent lock	112R	X1	Qualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified
2	W01814501801	Company 2	Smart door lock	H636	X2	Qualified	Unqualified	Qualified	Qualified	Unqualified	Qualified	Unqualified	Unqualified	Qualified
3	W01814501802	Company 3	Smart door lock	S400	X3	Qualified	Unqualified	Qualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified
4	W01814501803	Company 4	Smart fingerprint lock	VHS	X4	Qualified	Unqualified	Qualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified
5	W01814501804	Company 5	Cloud intelligent lock	HZ8-R-T	X5	Qualified	Unqualified	Qualified	Qualified	Unqualified	Unqualified	Unqualified	Unqualified	Unqualified
6	W01814501805	Company 6	Smart fingerprint lock	D2	X6	Qualified	Qualified	Qualified	Qualified	Qualified	Unqualified	Unqualified	Unqualified	Unqualified

Table 4: Risk grade classification matrix.

Risk factor possibility	The extent of the impact on the risk				
	Key	Serious	General	Small	Ignored
Very high	High	High	High	Medium	Medium
High	High	High	Medium	Medium	Low
Medium	High	Medium	Medium	Low	Low
Low	Medium	Medium	Low	Low	Low
Very low	Medium	Low	Low	Low	Low

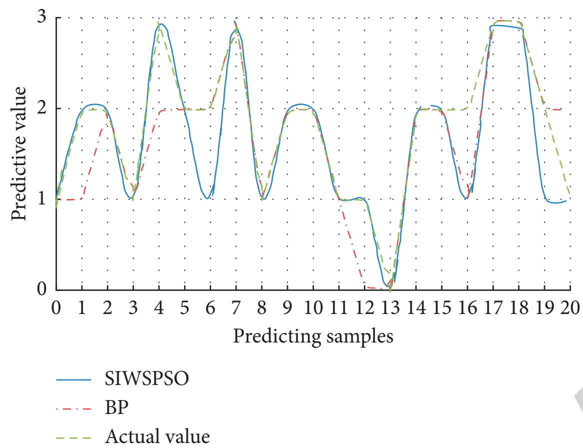


Figure 3: Comparison of training times between SIWSPSO-BP and BP.

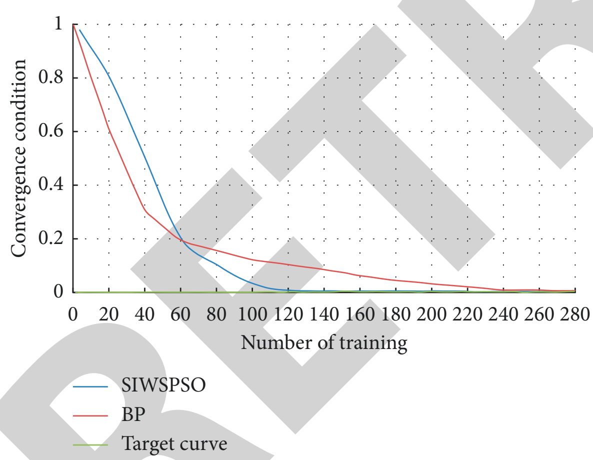


Figure 4: Comparison of prediction errors between SIWSPSO-BP and BP.

Therefore, product quality risk warning based on SIWSPSO-BP network can give higher warning to product risk.

5. Conclusions

In order to explore optimization method of BP network application and solve problems of BP neural network such as sensitivity to initial value, slow convergence rate, and local minimum point, this paper proposes the idea of optimizing BP network by PSO algorithm. In view of the disadvantages of the standard PSO algorithm, this paper innovatively proposes to use random variables to describe the inertia weight and adopts the strategy of asynchronous variation of learning factors, which is improved and optimized first. In order to verify and analyze the performance of SWISPSO-BP network, this paper establishes a product quality risk early warning model based on SIWSPSO-BP neural network and applies the optimized neural network algorithm to the quality risk early warning of product quality big data. Experiments show that the performance of SWISPSO-BP network has been obviously improved. The accuracy is increased by 10%, and the convergence speed is doubled. In the next step, we will collect more data used in scenarios and further study how to improve the performance of BP network.

Data Availability

The labeled datasets used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the Chongqing Creation Vocational College.

References

[1] E. Z. Farsa, A. Ahmadi, M. A. Maleki, M. Gholami, and H. N. Rad, "A low-cost high-speed neuromorphic hardware based on spiking neural network," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 9, pp. 1582–1586, 2019.

[2] Y. Koçak and G. Ü Şiray, "New activation functions for single layer feedforward neural network," *Expert Systems with Applications*, vol. 164, Article ID 113977, 2021.

[3] Y. J. Wong, S. K. Arumugasamy, C. H. Chung, and S. Anurita, "Comparative study of artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) for modeling of Cu (II) adsorption from aqueous solution using biochar derived from rambutan (Nephelium lappaceum) peel," *Environmental Monitoring and Assessment*, vol. 192, no. 7, pp. 1–20, 2020.

[4] P. M. Shakeel, A. Tolba, Z. Al-Makhadmeh, and M. M. Jaber, "Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks," *Neural Computing & Applications*, vol. 32, no. 3, pp. 777–790, 2020.

[5] W. Deng, R. Yao, H. Zhao, X. Yang, and G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm," *Soft Computing*, vol. 23, no. 7, pp. 2445–2462, 2019.

[6] B. Jamali, M. Rasekh, F. Jamadi, R. Gandomkar, and F. Makiabadi, "Using PSO-GA algorithm for training artificial neural network to forecast solar space heating system parameters," *Applied Thermal Engineering*, vol. 147, pp. 647–660, 2019.

[7] W. Li, Q. Pan, S. Liang, and J. Y. Jiao, "Research on fractal image compression hybrid algorithm based on convolutional

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9856892, 1 page
https://doi.org/10.1155/2023/9856892

Retraction

Retracted: Research on Prediction of News Public Opinion Guiding Power Based on Neural Network

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Chen and G. Li, "Research on Prediction of News Public Opinion Guiding Power Based on Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 2607492, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2607492, 9 pages
https://doi.org/10.1155/2022/2607492

Research Article

Research on Prediction of News Public Opinion Guiding Power Based on Neural Network

Shan Chen and Guang Li

School of Media Communications, Linyi University, Linyi, Shandong 276000, China

Correspondence should be addressed to Shan Chen; chenshan@lyu.edu.cn

Received 18 February 2022; Revised 7 March 2022; Accepted 22 March 2022; Published 18 April 2022

Academic Editor: Chin-Ling Chen

The development of big data technology and the popularity of we-media platforms make the prediction of news public opinion more complicated, which means that the complexity and dynamic nature of news public opinion require higher accuracy of prediction. With the rapid popularization of we media technology, traditional single-model algorithm is difficult to effectively predict network public opinion under the current background. Therefore, this paper proposes an algorithm based IGA-RBF neural network to deal with the complicated news public opinion prediction. Firstly, the ARMA (autoregressive moving average model) prediction model is constructed and the BRF neural network is combined. Then IGA is introduced to optimize BRF neural network, and the column vector of output matrix of hidden layer is optimized globally. The algorithm uses k-means clustering to select parameters in RBF network. The experimental results demonstrate that the model algorithm makes up for the shortcomings of the single prediction algorithm, improves the accuracy of prediction, and has better prediction results of public opinion trends.

1. Introduction

With the rapid development of Internet technology, China has entered a new era of information explosion, and users scattered on the network have become the creators of information [1]. Thus, the speed and breadth of information dissemination have spread exponentially with the booming development of online social media. While facilitating information dissemination, social individuals also add complexity to public opinion monitoring and prediction [2]. Internet and we-media users tend to be younger people who prefer to obtain news information through Weibo, WeChat, Douyin, and other means [3]. The information content released by traditional media is gradually ignored by people, and also, the authoritative information is less and less. The amount of information released through the Internet is so large that it is difficult to distinguish authenticity. News public opinion without legal constraints is prone to public opinion crisis [4]. Intense language caused by public opinion in the society tends to have a negative impact on the values and cognition of netizens, and negative news and public opinion increase the difficulty of social management [5]. When news time harms the image of the government, it will have a very negative effect on our government. Therefore, the study of the evolution of news public opinion is of great significance to the development of society [6]. New technologies and new platforms are used to quickly find news hotspots, grasp the development trend of news public opinion through prediction algorithms, and correctly guide news public opinion comments, so as to promote the healthy development of social and cultural industries [7].

With the continuous expansion of the scale of Internet users, the Internet has become another important way of information dissemination besides traditional media such as newspaper, radio, and TV [8]. At present, social media or websites relying on Internet platforms such as WeChat, Tiktok, Weibo, forums, Post bar, and news websites have become important battlefields of information dissemination [9]. An increasing number of netizens receive information from the real world and the virtual world and express their views and opinions on social events freely on social media. When sudden network events or social events attract a large number of netizens' attention, they evolve into online public opinions, which will have a significant impact on social public security and long-term development [10]. In the era of big data, a huge amount of data needs to be collected for information work. Meanwhile, social media, news media, search engines, and other network platforms make network communication present a "honeycomb" divergent structure, making it more difficult to analyze and predict network and situation [11]. In addition, the warning time is unstable with the outbreak of emotional events. These problems make the response to online public opinions face greater challenges.

Literature [12] first clearly defined the negative public opinions of the government by taking microblog as the information carrier. Then it establishes the government negative public opinion prediction star based on Markov chain, which provides theoretical support for the government to deal with negative public opinion timely and reasonably and guide the trend of public opinion. Lyapunov was used in literature [13] to strongly prove that the trend development of network has chaotic characteristics, and related data were reconstructed in phase space to prove the practicability of the algorithm. Literature [14] improved the accuracy of microblog public opinion prediction by combining improved SEIR model and PageRank algorithm with Bayesian network. Literature [15] proposed an emotion perception time series prediction method based on dynamic time warping and autoregressive comprehensive moving average model to predict topic popularity. Literature [16] innovatively constructed Baidu Index time series index of network events and adopted relevant time data for training. Literature [17] studies the prediction and discovery of Weibo hot topics from the perspectives of data mining and term terms, respectively. Literature [18] studies the changing trend of public emotion based on the emotional common sense of microblog time. Literature [19] analyzes the mechanism and trend prediction of Wang Min's mood changes under the background of public opinion big data.

Many algorithms still need to adjust a large number of parameters in the optimization process, which increases the complexity of adjusting parameters in the process of network training, resulting in more network iterations and slower convergence. Aiming at the above problems, this paper proposes a research algorithm of news public opinion prediction based on improved RBF neural network. The innovations and contributions of this paper are listed below.

(1) The algorithm uses k-means clustering to select parameters in RBF network.
(2) AR-RBF combined prediction model is constructed, and historical data are selected as training samples of RBF neural network and ARMA time series.
(3) The normalized order of RBF neural network was optimized by IGA. On the premise of satisfying the target error, the column vector of the output matrix of the hidden layer of the network is optimized.
(4) The crossover probability and mutation probability of genetic calculation are adjusted adaptively according to the current individual fitness and the number of evolutionary iterations, aiming to find the optimal orthogonalization order, reduce the average error, and improve the accuracy of prediction.

The chapter structure of this paper is as follows. The related theory is described in Section 2. The recommendation method is constructed in Section 3. Section 4 focuses on the experiment and analysis. Then Section 5 is the conclusion.

2. The Related Theory

2.1. RBF Neural Network. Radial basis function (RBF) [20] neural network does not fall into local extremum and can approximate any nonlinear function with arbitrary precision. Therefore, RBF neural network is chosen as the algorithm of public opinion prediction. RBF neural network is composed of input layer, hidden layer, and output layer. Its structure is shown in Figure 1.

In this paper, the commonly used Gauss function is chosen as the radial basis function of RBF neural network.

$$R_y(I) = \exp\left(-\frac{\left\|I - C_y\right\|2}{2\sigma_y^2}\right), \quad (1)$$

where y is the number of nodes of the hidden layer. $R_y(I)$ is the output of the node at the hidden layer y of the network. I is the input vector. C_y is the center of the y radial basis function. σ_y is the scale factor of the y implicit node, which determines the radial action width of the function. Then through continuous learning, radial basis algorithm is used to adjust the weight ω_{yw} from the y implicit node to the w output node to minimize the learning error. Finally, the connection weight is fixed and the corresponding output is obtained from the input of the network.

2.2. Parameter Selection Based on K-Means Clustering. Because the number of hidden layer nodes y of RBF neural network is too small, it will lead to high error. Increasing the number of hidden layer nodes y can reduce the error of training. But too much increasing will affect the generalization ability of neural network and increase the network complexity, resulting in slow training speed. In principle, the minimum number of hidden layer nodes should be selected to meet the accuracy requirements. The key parameters of hidden layer are center C_y and width σ_y of hidden layer node. The center of traditional RBF neural network is the sample. Its width is determined by the number of data centers and the maximum distance between them and cannot be dynamically optimized. Therefore, it is necessary to determine the proper center C_y and width σ_y of the hidden layer to improve the performance of RBF neural network.

K-means clustering algorithm is a common unsupervised clustering algorithm with fast training speed and good ductility. Therefore, k-means clustering algorithm is used to determine the center C_y and width σ_y of the hidden layer to optimize the RBF neural network. The initial category of k-means clustering algorithm samples is unknown, so the number of clustering centers needs to be set. In RBF neural network, the center C of Gaussian function is the clustering center. After determining the initial cluster center, formula (2) is used to calculate the distance between the input layer data and the cluster center:

$$d_t(y) = \left\| i_t - C_y \right\|, \tag{2}$$

where i_t represents the n input vector. C_y represents the y cluster center. $d_t(y)$ is the distance from i_t to C_y.

According to the minimum distance between the sample point and each cluster center, the data is divided into K pieces. The mean of the distance between all samples in each cluster is the new cluster center C'_y.

$$C'_y = \frac{1}{T_y} \sum_{P_y} I, \tag{3}$$

where P_y is the sample set of the y cluster center. T_y is the amount of data in the set. When the new clustering center is different from the initial clustering center, the above steps are repeated until the clustering center is fixed and the hidden layer node center C_y is obtained. The distance between node centers of each hidden layer is calculated and the minimum value is taken as the scale factor σ_y.

$$\sigma_y = \lambda \min_x \left\| C_y - qC_x \right\|. \tag{4}$$

In the formula, λ is the overlap coefficient, which is generally 1 initially and then adjusted through experiments.

Traditional k-means clustering algorithm requires K initial clustering centers. The minimum number of hidden layer nodes to reach the target precision is the number of hidden layer nodes of RBF neural network; then the center C_y and the width σ_y of hidden layer node are determined.

3. The Proposed Method

3.1. Linear Programming Determines Combination Weights. The standard model of general linear programming is shown in

$$\max k = \sum_{y=1}^{t} c_y i_y. \tag{5}$$

$$\text{s.t.} \begin{cases} \sum_{y=1}^{t} g_{xy} i_y = h_x & x = 1, 2, \cdots, w \\ i_y \geqslant 0 & y = 1, 2, \cdots, t \end{cases} \tag{6}$$

If solution $i = (i_1, i_2, \ldots, i_t)$ satisfies (6), $i = (i_1, i_2, \ldots, i_t)$ is said to be feasible. If $i = (i_1, i_2, \ldots, i_t)$ simultaneously satisfies (5) to obtain the maximum value, it is called the optimal solution.

3.2. ARMA Model. Assume that the sequence $\{I_n, n = 0, \pm 1, \pm 2, \cdots\}$ satisfies the following model:

$$I_n - \phi_1 I_{n-1} - \cdots - \phi_u I_{n-u} = \varepsilon_n - \theta_1 \varepsilon_{n-1} - \cdots - \theta_v \varepsilon_{n-v}. \tag{7}$$

I_n is zero mean stationary sequence. ε_n is stationary white noise with zero mean and variance σ_x^2. I_x is the ARMA sequence, whose order is u and v. Applying the two operator polynomials $\phi(H)$ and $\theta(H)$, (7) can be written as $\phi(H) I_x = \theta(H) \varepsilon_n$.

For a general stationary sequence $\{I_n, n = 0, \pm 1, \pm 2, \cdots\}$, set its average value $E(I_n) = \mu$. It satisfies the following model:

$$(I_n - \mu) - \phi_l (I_{n-1} - \mu) - \cdots - \phi_u (I_{n-u} - \mu) = \varepsilon_n - \theta_1 \varepsilon_{n-1} - \cdots - \theta_v \varepsilon_{n-v}. \tag{8}$$

If the operator polynomials $\phi(H)$ and $\theta(H)$ are used, (8) can be expressed as $\phi(H)(I_x - \mu) = \theta(H)\varepsilon_n$.

The steps for building the ARMA model are as follows.

3.2.1. AIC Criterion for ARMA Model Grading. Akaike information criterion (AIC) [21] determines the order of quasiside by selecting u and v, which satisfies the following equation:

$$\min AIC = t \ln \hat{\sigma}_s^2 + 2(u + v + 1). \tag{9}$$

In (9), t represents the capacity of the selected sample. $\hat{\sigma}_s^2$ represents the expected value of variance σ_s, and the stationarity of time series is related to u and v. If AIC has the minimum value when $u = \hat{u}$, $v = \hat{v}$, sequence $ARMA(u, v)$ is considered to have the best stationarity.

When sequence $ARMA(u, v)$ contains unknown mean value parameter μ, the model is $\phi(H)(I_n - \mu) = \theta(H)\varepsilon_n$. In this case, the number of unknown parameters is $z = u + v + 2$. AIC criterion is expressed where u and v are selected to satisfy

$$\min AIC = n ln \hat{\sigma}_\varepsilon^2 + 2(u + v + 2). \tag{10}$$

In fact, the minimum value of (10) above is the same as that of (10) above. When the values of (u, v) are set as $(\hat{u}, \hat{v})$, the values of (9) and (10) are the smallest.

3.2.2. χ^2 Test of ARMA Model. For sequence AR(u), assuming $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_u$ is the unknown parameter estimate, then the residual $\hat{\varepsilon}_n = X_n - \hat{\phi}_1 X_{n-1} - \cdots - \hat{\phi}_u X_{n-u}$, $n = 1, 2, \cdots, t$ (assuming $I_0 = I_{-1} = \ldots = I_{1-u} = 0$) is denoted by

$$\eta_z = \frac{\sum_{n=1}^{t-z} \hat{\varepsilon}_n \hat{\varepsilon}_{n+z}}{\sum_{n=1}^{z} \hat{\varepsilon}_n^2}, \quad z = 1, 2 \cdots, L. \tag{11}$$

where L is the mantissa of the autocorrelation function of $\hat{\varepsilon}_n$, and Ljung's χ^2 test statistic is

$$\chi^2 = t(t+2) \sum_{z=1}^{L} \frac{\eta_z^2}{t - z}. \tag{12}$$

3.3. The Improved Genetic Algorithm. The traditional genetic algorithm (GA) has a strong global search ability for all possible solution sets, though the local search ability is weak. As a result, the convergence rate becomes slow and it is difficult to approach the global optimal solution. In addition, traditional genetic algorithms only judge the merits and demerits of solutions

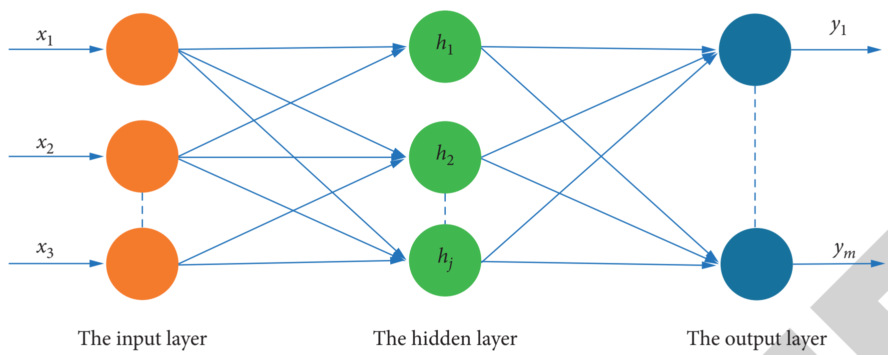


Figure 1: RBF neural network structure.

according to fitness. Consequently, in the early stage of evolution, some individuals' fitness value is too large, which leads to the population falling into the local optimal solution, resulting in premature convergence. Therefore, the crossover operator combined with genetic algorithm can change the global search ability, and the mutation operator can change the local search ability. Besides, the crossover probability and mutation probability of genetic operator can be adjusted adaptively according to the current individual fitness and the number of evolutionary iterations.

In traditional genetic algorithms, the crossover probability U_c is fixed. Generally, the maximum crossover probability U_{cmax} is not more than 0.8, and the minimum crossover probability U_{cmin} is not less than 0.3. If the selected individuals have a large crossover probability, the search scope of genetic algorithm will be expanded, and the global search ability will be enhanced. However, excessive crossover probability may lead to the destruction of the original high adaptability of chromosomes. If the crossover probability of selected individuals is small, the global search ability of genetic algorithm will be reduced and the convergence speed will be slow. Therefore, in the process of evolution, it is necessary to continuously adjust the crossover probability according to the current individual fitness and the number of evolutionary iterations.

In the early stages of evolution, individuals are usually less able to adapt to their environment. Therefore, when the individual fitness value is lower than the average fitness value, a larger crossover probability should be selected to increase the global search range of the algorithm. In the later stage of evolution, with the increasing number of iterations, individuals with low environmental adaptability have been gradually eliminated. Hence, when the individual fitness value is higher than the average fitness value, the crossover probability can be appropriately reduced according to the number of iterations, thus reducing the global search ability of the algorithm. Based on the above ideas, the improved adaptive crossover probability is

$$U_c' = \begin{cases} U_{cmax}, & F_{\max} < F_{\mathrm{mcan}} \\ U_{cmax} - \dfrac{U_{cmax} - U_{cmin}}{iter_{\max}} \times iter, & F_{\max} \geqslant F_{\mathrm{mean}} \end{cases}, \quad (13)$$

where $iter$ and $iter_{\max}$ represent the current iteration number and maximum iteration number of the algorithm, respectively, $F_{\max}$ represents the maximum fitness value of the two individuals to be crossed in the parent population, and F_{mean} represents the average fitness value of all individuals in the parent population. The fitness F was determined using the distorted form of the error function E between the actual output and the target output of the IGA-RBF network. The fitness of the F individual can be expressed as

$$F(i) = \frac{1}{E(i)}. \quad (14)$$

In traditional genetic algorithm, mutation probability U_w is fixed. Generally, the maximum variation probability U_{wmax} is not more than 0.1, and the minimum variation probability U_{wmin} is not less than 0.001. If the mutation probability is small, some important genes of the good chromosomes produced by crossover operation can be retained. However, if the mutation probability is large, the algorithm will always be in the state of random search, making it difficult for the excellent individuals produced by crossover operation to continue genetic operation. Therefore, in the process of evolution, it is necessary to continuously adjust the mutation probability according to the current individual fitness and the number of evolutionary iterations.

Similar to the crossover process, in the early stage of evolution, when the fitness value of an individual is lower than the average fitness value, small mutation probability is selected to preserve the excellent genes in the chromosome as much as possible. In the later stage of evolution, when the individual fitness value is higher than the average fitness value, the mutation probability can be appropriately increased according to the number of iterations, so as to improve the local search ability of the algorithm. Therefore, the adaptive mutation probability is put forward as

$$U_w' = \begin{cases} U_{wmin}, & F < F_{\mathrm{mean}} \\ U_{wmin} + \dfrac{U_{wmax} - U_{wmin}}{iter_{\max}} \times iter, & F \geqslant F_{\mathrm{mean}} \end{cases}, \quad (15)$$

where $iter$ and $iter_{\max}$ represent the current iteration number and maximum iteration number of the algorithm, respectively, F represents the fitness value of chromosomes to be mutated in the parent population, and F_{mean} represents the average fitness value of all chromosomes in the parent population.

Since it is not guaranteed that the new individuals, which will be generated after crossover or mutation, have better environmental adaptability than the parent individuals, it is necessary to compare the fitness of the parent individuals and offspring individuals. Among them, the paternal individual represents the chromosome before crossover or mutation, and the offspring individual represents the chromosome after crossover or mutation. If the fitness value of the offspring is greater than that of the parent, it will indicate that the fitness of the offspring increases after crossover or mutation, and the offspring can continue to be genetically operated. On the contrary, if the fitness of the individual decreases after crossover or mutation, the offspring will be discarded and the parent will be used to replace the offspring to continue the genetic operation.

3.4. The Optimized RBF Neural Network with IGA. It is, in the process of training and optimization of RBF neural network, mainly to learn and adjust the center of hidden layer activation function, expansion constant, and continuous weight from hidden layer to output layer. For solving these network parameters, the learning algorithm of the network is different because of the different determination methods. Therefore, there are various learning algorithms of RBF neural network, such as self-organizing center selection method and orthogonal least square method. According to the different methods of selecting the center of the activation function of the hidden layer, the learning algorithm of RBF neural network can be divided into two types. (1) Input samples are used as the center of the activation function, which will not change once it is determined, such as orthogonal least square method. (2) The activation function center is not fixed, but can be dynamically adjusted continuously through RBF neural network training, such as self-organizing center selection method. Since the self-organizing center selection method needs to determine the number of hidden layer nodes manually, the number of hidden layer nodes has a direct influence on the determination of activation function center and expansion constant. The unreasonable number of hidden layer nodes may lead to the deterioration of RBF neural network performance, nonlinear approximation ability, and generalization ability. Therefore, this paper chooses orthogonal least square method, which is widely used and can automatically design a minimum network structure, as the learning algorithm of RBF neural network.

In order to reduce the complexity of adjusting parameters during network training, IGA was used to optimize the orthogonalization sequence of RBF neural network based on orthogonal least square method. Orthogonal least square method orthogonalizes the column vectors of the output matrix of the hidden layer through Gram–Schmidt [22] orthogonalization, so as to find the column vectors that contribute the most energy to the output of the network. Although the difference in the order of orthogonalization will lead to the change of the column vectors that contribute the most energy to the output, the order of selecting the hidden layer activation function centers from the input samples will also change. However, the total training error of RBF neural network does not change because of its different order. Therefore, the improved genetic algorithm can be used to optimize the column vector of output matrix of hidden layer of RBF neural network on the premise of satisfying the target error. According to the optimal column vector, the optimal orthogonalization sequence is determined. Finally, the optimal activation function center and hidden layer node number are obtained, and the IGA-RBF neural network with better structure is designed, so as to reduce the tedious degree of adjusting parameters in network training. Specific steps using IGA optimization based on orthogonal least squares are shown below.

Step 1: initialize the network. All input sample matrices $I = [i_1, i_2, \cdots, i_t]$ were taken as the center of the activation function, and the number of nodes in the hidden layer was initialized to t. At the same time, set the maximum iteration number l_{max} of the genetic algorithm, and set the initial iteration number l to 0. When an algorithm iteration is completed, the value of l automatically increases by 1.

Step 2: compute the output matrix $\Phi = [\varphi_1, \varphi_2, \cdots, \varphi_t]$ of the hidden layer, where $\varphi_x = \varphi(I, c_x)$ is the output vector of the x hidden layer element. When training sample i_z is input, the output of the x hidden layer unit is $\varphi(i_z, c_x)$, also known as the activation function of the hidden layer. It is usually expressed by Gaussian function, which is expressed by

$$\varphi(i_z, C_x) = \exp\left(\frac{-\|i - c_x\|^2}{2\sigma_x^2}\right), \tag{16}$$

where c_x is the center of the activation function and σ_x is the expansion constant.

Step 3: since the problem to be solved is to find the optimal order of t column vectors of Φ, by sorting t different column vectors, $t!$ possible solutions can be obtained to form the solution space of the problem in question. Genetic algorithm can not directly optimize the solution space data to solve the problem, so it needs to map it into the genetic space by means of gene coding. First, the different sequences of t column vectors of Φ are numbered successively, and then these numbers are processed by binary encoding so that they can represent the solution of the problem to be solved in the form of 0/1 strings.

Step 4: initialize the population. From the solution space of the problem to be solved, the possible solution sets of groups μ were randomly selected to form the initial parent population $U_0 = [u_1, u_2, \cdots, u_\mu]$, where $u_x\,(x = 1, 2, \ldots \mu)$ is a set of possible solutions to the problem, as well as a set of orthogonalized sequences.

Step 5: the negative gradient steepest descent method is used to continuously modify the connection weight ω_{xy} between the hidden layer and the output layer. When the total error E of the network is less than the target error ε, the correction is stopped.

Step 6: use the genetic operator.

Step 6.1: to calculate the individual fitness value, according to the individual fitness value and roulette selection criteria, two better individuals u_x and v_y were selected from the parent population.

Step 6.2: cross u_x and v_y according to the improved crossover probability U_c' to generate daughter chromosomes.

Step 6.3: mutate the daughter chromosomes generated in Step 6.2 according to the improved mutation probability U_w'.

Step 7: when the iteration number l of the algorithm exceeds the set maximum iteration number, the genetic evolution process is planted. In this case, you can determine the number of nodes at the hidden layer. Otherwise, proceed to Step 5 to continue the next generation of genetic operations.

Step 8: if the termination conditions of the genetic algorithm are met, the chromosome with higher fitness value can be selected as the most individual in the solution space through the calculation of fitness function. At this point, the center of the network activation function can be determined according to the most individual obtained, namely the optimal orthogonalization order. As the expression form of this individual is genotype string structure, decoding operation is required, that is, the inverse operation of binary coding.

3.5. Linear Optimization Method to Determine the Weight. The key steps of AR-RBF combined prediction model are as follows.

(1) Obtain the required data, given the training data set of AR-RBF model, including the sample set of input data and output data.

(2) The feature values are obtained by feature extraction, standardization, and normalization of the original data of shooting topics.

(3) Calculate the absolute value of relative prediction error $re_{xy} = |(j_{xy} - n_y)/n_y|$ of each unit model prediction method (ARMA time series model and improved RBF neural network prediction model),

where re_{xy} is the absolute value of error of variable y in model x. n_y is the expected output of the y variable. j_{xy} is the actual output value of the y variable in the x model. $x = 1, 2, \ldots, w$, and $y = 1, 2, \ldots, w$ are the number of prediction models, and T is the number of prediction model variables.

(4) Linear optimal solution is used to determine the weighting coefficient of AR-RBF model. Such that the weighting coefficient $i_x\,(x = 1, 2, \ldots, w)$ satisfies $\min k = \sum_{y-1}^{W} |\sum_{x=1}^{w} i_x j_{xy} - n_y|$, where k is the total value of prediction error of the model.

(5) After the weighted factor i_x is calculated, the weighted factor is used to recalculate the predicted value $\hat{b}$ of the RBF combined prediction model $\hat{b} = \sum_{x=1}^{w} i_x j_{xy}$.

4. Experimental Results and Analysis

The experimental environment of this algorithm is Win10 system, with Intel dual-core processor, 8G memory, and 500G hard disk. The data were obtained from Weibo and Baidu, and the Terminal High Altitude Area Defense (THAAD) incident in South Korea was taken as a hot topic. Also, the respective Baidu index and Weibo index are obtained. The duration of the relationship was 1 month.

4.1. The Effectiveness of the Proposed Algorithm. The RBF neural network and the proposed algorithm are selected to compare their prediction accuracy, so as to verify the prediction effect of the proposed algorithm. Figure 2 shows the comparison of RBF neural network with the predicted click-through number calculated in this paper and the actual click-through number of Weibo. R square value, $RMSE$ value, and MSE value were used as evaluation indexes to measure goodness of fit between predicted value and actual value. The equation is shown in the following equation:

$$R^2 = 1 - \frac{\sum_x \left(\hat{j}_x - j_x\right)^2}{\sum_x \left(\hat{j}_x - j_x\right)^2},$$

$$RMSE = \sqrt{\frac{1}{w}\sum_{x=1}^{w}\left(j_x - \hat{j}_x\right)^2}, \tag{17}$$

$$MSE = \frac{1}{w}\sum_{x=1}^{w}\left(j_x - \hat{j}_x\right)^2,$$

where w is the total number of samples. j is the true value, and $\hat{j}$ is the predicted value. $RMSE$ is the root mean square error and MSE is the mean absolute error.

The number of microblog clicks is a direct reflection of the changing barometer of sentiment. As can be seen from Figure 2, the evolution of the heat wave is characterized by rapid outbreak and slow decline, experiencing four stages of germination, acceleration, maturity, and decline, basically in line with the life cycle of online public opinion. To be specific, BP neural network in the stage of budding and the prediction results are not accurate and the delay is serious, especially in the outbreak of love and its predicted value deviates from the real value are large. As for the algorithm in this paper, the predicted trend not only is consistent with the real situation in a relatively flat place, but also can accurately reveal the trend of the real value when the wave peak is large, which shows, that based on the improved RBF neural network algorithm, we can mine the internal law of love communication from the increasing media data. At the same time, the development trend of the situation is predicted accurately, and the error of the predicted result is smaller

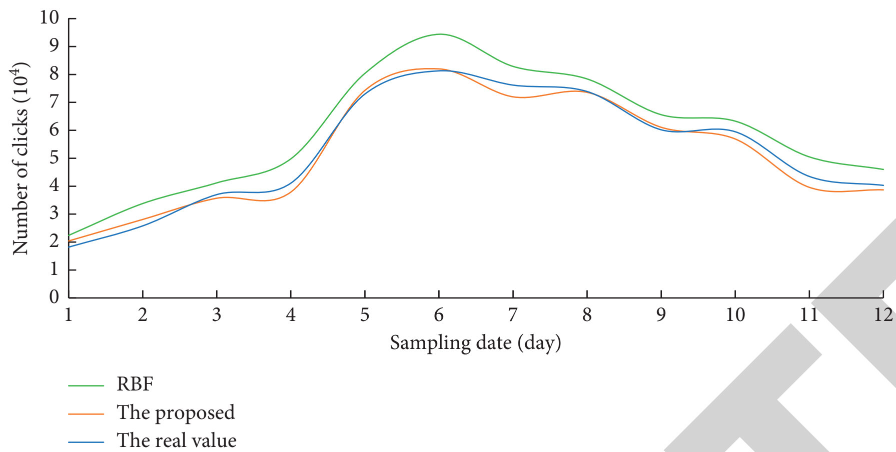


Figure 2: The predicted results with RBF neural network.

Table 1: Prediction results of the proposed algorithm and RBF neural network.

Methods	R^2	RMSE	MSE
RBF	0.8336	0.0286	0.00075
The proposed	0.9361	0.0148	0.00018

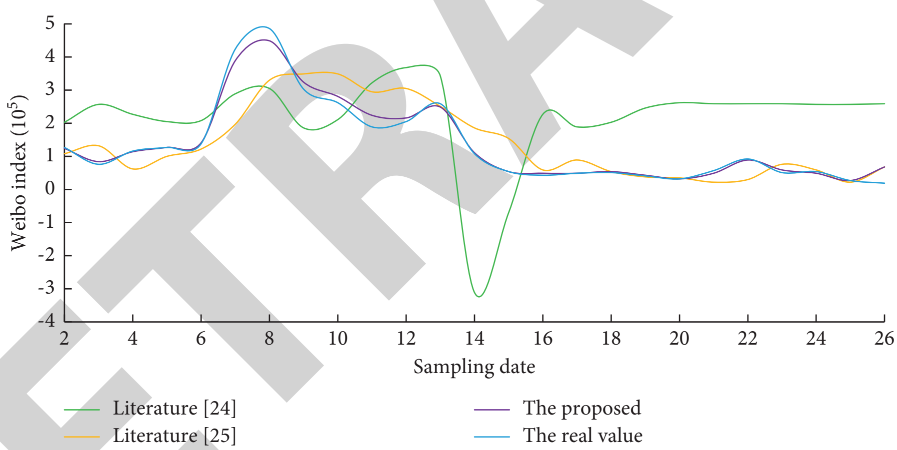


Figure 3: The predicted results of THAAD incident on Weibo index.

than the actual value. The R square value, RMSE value, and MSE value of the prediction results of BPF neural network algorithm and the algorithm in this paper are shown in Table 1.

Comparative analysis of the data in Table 1 shows that each index of the proposed algorithm is significantly better than that of RBF neural network algorithm. Its R square value is 0.9361, higher than the BRF neural network algorithm 0.8336. Compared with BRF neural network, the RMSE value of the proposed method is decreased by 48.2%. MSE value is greatly reduced compared with BRF neural network. Experimental results display that the proposed algorithm has higher prediction accuracy and is an effective method for public opinion trend prediction.

4.2. The Comparative Analysis with Other Algorithms. Three algorithms are adopted to predict the Weibo index of the THAAD incident, as shown in Figure 3.

As can be seen from Figure 4, the fitting effect of literature [23] is the worst, and the prediction results obtained cannot be well fitted to public opinion data. There is a big gap between the sum effect of literature [24] and the trend of public opinion, and the trend effect of public opinion

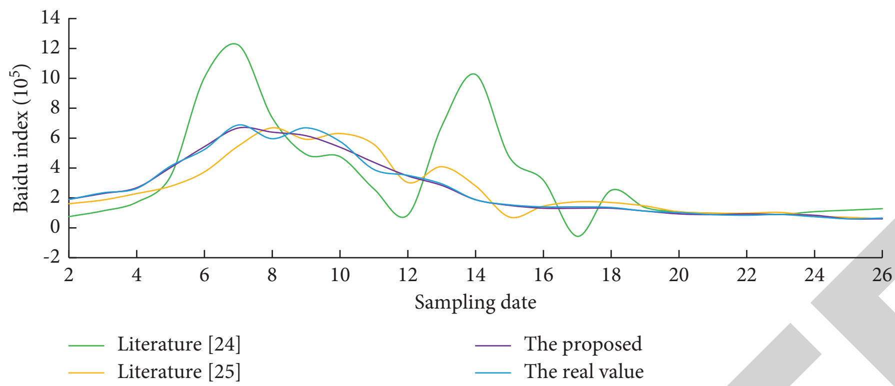


Figure 4: The predicted results of THAAD incident on Baidu index.

obtained is average. The residual error fitting effect of the algorithm in this paper is good and the trend of public opinion is close to the actual trend of public opinion.

Three algorithms are used to predict the Baidu index of the THAAD incident, as shown in Figure 4.

As can be seen from Figure 4, the fitting effect of literature [23] is poor, and it is unable to control the change rule of the trend of and situation well, and the predicted results are unstable. The prediction results obtained by literature [25] and the algorithm in this paper are better, but the prediction results obtained by literature [24] are not as good as the prediction results of the algorithm in this paper. The prediction results of the target prediction value of the algorithm in this paper are the most accurate among the three algorithms.

5. Conclusion

With the rapid development of mobile communication technology and Internet technology, social hot issues have been rapidly spread and entered the public view. Consequently, network public opinion data becomes increasingly complex, while detailed public opinion analysis and accurate prediction of public opinion development are of particular importance. At present, the prediction research of news public opinion is still in the development stage. In addition, there are still a great number of problems to be solved. This paper proposes an algorithm of news public opinion prediction based on IGA-RBF neural network. The algorithm uses Weibo index and Baidu index as news public opinion data sources to study the news public opinion trend of "THAAD incident." Firstly, the improved BRF neural network and ARMA time series are trained, and the weight of AR-RBF combined prediction model is determined by one linear optimal method. The experimental results indicate that the proposed algorithm not only has a good fitting effect, but also has a small deviation. The target prediction value of the proposed algorithm is more accurate, providing more reliable data for news public opinion prediction. In the future, the author will conduct prediction experiments on more news platforms and make a more systematic analysis of the performance of the algorithm in this paper.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Ministry of Education Humanities and Social Sciences Research Youth Foundation "Research on Improving the Public Opinion Guidance of Party Media Based on Big Data" (Project No. 17YJCZH023).

References

[1] L. Liu, S. Yu, X. Wei, and Z. Ning, "An improved Apriori-based algorithm for friends recommendation in microblog," *International Journal of Communication Systems*, vol. 31, no. 2, p. e3453, 2018.
[2] M. Lim, "Freedom to hate: social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia," *Critical Asian Studies*, vol. 49, no. 3, pp. 411–427, 2017.
[3] S. Yang, *We-media as a Marketing Tool for Fashion Business in China: Case*, LingSiYaoYao Ltd, Chennai, Tamilnadu, 2020.
[4] C. Zhou, F. Su, T. Pei et al., "COVID-19: challenges to GIS with big data," *Geography and sustainability*, vol. 1, no. 1, pp. 77–87, 2020.
[5] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time series," in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social media*, Washington, DC USA, May 2010.
[6] A. Hermida, F. Fletcher, D. Korell, and D. Logan, "Share, like, recommend: decoding the social media news consumer[J]," *Journalism Studies*, vol. 13, no. 5-6, pp. 815–824, 2012.
[7] Q. Jia, Y. Guo, G. Wang, and S. J. Barnes, "Big data analytics in the fight against major public health incidents (including COVID-19): a conceptual framework," *International Journal of Environmental Research and Public Health*, vol. 17, no. 17, p. 6161, 2020.
[8] M. Kaigo, "Social media usage during disasters and social capital: twitter and the Great East Japan earthquake[J]," *Keio communication review*, vol. 34, no. 1, pp. 19–35, 2012.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9865913, 1 page
https://doi.org/10.1155/2023/9865913

Retraction

Retracted: Application of Hidden Markov Model in Financial Time Series Data

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Chang and J. Hu, "Application of Hidden Markov Model in Financial Time Series Data," *Security and Communication Networks*, vol. 2022, Article ID 1465216, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1465216, 10 pages
https://doi.org/10.1155/2022/1465216

Research Article

Application of Hidden Markov Model in Financial Time Series Data

Qingqing Chang and Jincheng Hu

School of Information Management, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

Correspondence should be addressed to Jincheng Hu; 20119876@lixin.edu.cn

Received 9 January 2022; Revised 26 February 2022; Accepted 19 March 2022; Published 16 April 2022

Academic Editor: Chin-Ling Chen

Financial time series have typical characteristics such as outliers, trends, and mean reversion. The existence of outliers will affect the effectiveness of the unknown parameter estimation in the financial time series forecasting model, so that the forecasting error of the model will be larger. Quantitative forecasting methods are divided into causal forecasting method and time series forecasting method. The causal forecasting method uses the causal relationship between the predictor variable and other variables to predict, and the time series forecasting method infers the future value of the predictor variable based on the structure of the historical data of the predictor. Therefore, this paper proposes a hidden Markov model prediction method based on the observation vector sequence, which can simultaneously consider the influence of the variable sequence structure and related factors.

1. Introduction

The financial market can be said to be the core of a country's economic operation. Governments and investment institutions in various countries strive to study the law of changes in the financial market, effectively manage it, and increase the rate of return on financial investment. Financial time series is the most important type of data in the economic and financial fields. Analysis, prediction, and control of this type of data are important tasks for financial workers and researchers. Financial time series include bonds, exchange rates, interest rates, stock prices, and financial futures prices. From a macroperspective, they can also include macroeconomic data such as investment, income, and consumption. The securities market plays an important role in the financial market and is the main body of the financial market. Especially in the context of increasing global financial integration, the degree of integration between our country's economy and the world economy is getting higher and higher, and the financial industry is facing greater and newer development opportunities and challenges. Understanding and grasping the essential laws of the financial market is directly related to the stability, efficiency, and safety of our country's financial market. After decades of rapid development, our country's financial industry has formed a considerable scale. The financial market is a complex system that will be affected by various factors. The law of its motion is difficult to predict, while the financial time series data are positive. It is the external manifestation of this characteristic. "Phenomena reflect the essence, and the essence determines the phenomenon." Therefore, there must be a lot of information about economic laws hidden in the financial time series. Finding all kinds of valuable information from financial time series and better understanding its application is undoubtedly of great significance to investment forecasting, decision making, and risk management [1–9].

In the 1960s, mankind entered a new era in the development of time series analysis theories and methods. Someone proposed the maximum entropy spectrum (MSE) estimation theory since the power spectrum estimation and the maximum entropy spectrum estimation effect of the autoregressive model is the same and it is later called modern spectrum estimation. Due to its inherent advantages, it has expanded the application field of time series analysis methods and thus has been more widely used. It has also made a great contribution to the research in the engineering field and has become a more practical tool [10–15].

From 1970 to 1979, information signal technology developed rapidly. In terms of theory, time series analysis methods have become more and more perfect. Improvements in theory have made it perform in some aspects such as parameter estimation algorithms, order determination methods, and modeling. Especially, we see that time series analysis methods are becoming more and more practical, and then some practical time series analysis software has appeared. Gradually, time series analysis methods and software have become impossible to predict and analyze those seemingly unrelated data [16–23].

From market forecasts to output forecasts and from population forecasts to earthquake forecasts, the theoretical development of time series analysis is becoming more and more mature and perfect. Nowadays, there are many research studies in the fields of precision instrument measurement and control, safety and quality inspection, maintenance, and measurement and control. Innovative discoveries provide convenient conditions for testing and management in the engineering and safety fields, so as to better ensure the precision of instruments and the accuracy of construction projects and ensure that high standards and high requirements for production safety and production quality are met [24–29].

Forecasting methods are divided into qualitative forecasting methods and quantitative forecasting methods, among which quantitative forecasting methods are divided into causal forecasting methods and time series forecasting methods. The causal prediction method uses the causal relationship between the predictor variable and other variables to make predictions. The correct determination of the causal relationship is the key to this type of method. The specific methods are univariate linear regression, multiple linear regression, and nonlinear regression prediction methods; the principle is to infer the future value of the predictive variable's historical data structure. Correctly identifying the historical data structure is the key to this type of method. Specifically, there are time series decomposition analysis method, moving average method, exponential smoothing method, ARIMA model method, gray forecasting method, Kalman filtering method, neural network prediction method, etc. Any time series as the target of prediction has many characteristics, such as the structure of the sequence itself, the causal relationship between it and other time series, and so on. The sum of these characteristics determines the future development of the time series. Only by reflecting these characteristics to a certain extent can a better forecasting effect be produced. However, no matter what kind of realistic prediction method, it can only describe a certain aspect of the prediction object but cannot describe all its characteristics. For example, time series can only describe the structure of the variable's own sequence but cannot describe other characteristics; the causal prediction method can describe the causal relationship between a variable and other variables, but it lacks the function of describing the structure of the variable's own time series. Bates and Granger first proposed the combined forecasting method in 1969. Its essence is to obtain new forecasting results from the forecasting results of several different forecasting methods according to a certain mapping. The combined forecasting method can simultaneously consider the structure of the sequence itself and its influencing factors, thereby effectively improving the prediction accuracy. The literature has conducted research on combined prediction, but it is not easy to determine the mapping relationship.

The hidden Markov model is a kind of Markov chain, and its state cannot be directly observed but can be observed through observation vector sequence; each observation vector is expressed as various states through some probability density distribution, and each observation vector is generated from a sequence of states with a corresponding probability density distribution. The hidden Markov model (HMM) originated a long time ago and was successfully applied to acoustic signal modeling in the 1980s, as shown in Figure 1. At present, HMM is mainly used in engineering fields, such as image processing, artificial speech synthesis, seismic exploration, biological signal processing, and so on, and has achieved important results with scientific significance and application value. In recent years, there are also documents that apply HMM to the volatility analysis of financial markets.

Therefore, the hidden Markov model is a double random process—a hidden Markov chain with a certain number of states and a set of explicit random functions. Firstly, introduce the basic theory of the hidden Markov model; secondly, on the basis of model training and hidden state sequence estimation, propose a hidden Markov model prediction algorithm based on observation vector sequence; finally, conduct simulation experiments and empirical research to prove the effectiveness of the method. The research results show that the hidden Markov model proposed in this paper adopts the mechanism conversion hybrid forecasting model and applies it to the forecast of financial time series, which effectively reduces the forecast error.

2. Basic Theory of Hidden Markov Model

The hidden Markov model is a double-embedded random process, that is, the hidden Markov model is composed of two random processes, one is an implicit state transition sequence, which corresponds to a simple Markov process (shown in Figure 2); the other is related to the implicit state transition sequence.Of these two random processes, one of the random processes (hidden state transition sequence) is unobservable and can only be inferred from the output observation sequence of the other random process, so it is called a hidden Markov model. Its basic elements include the following:

(1) A collection of hidden state numbers. The discrete set S is often used to represent different hidden states:

$$S = \{S_1, \ldots, S_N\}, \tag{1}$$

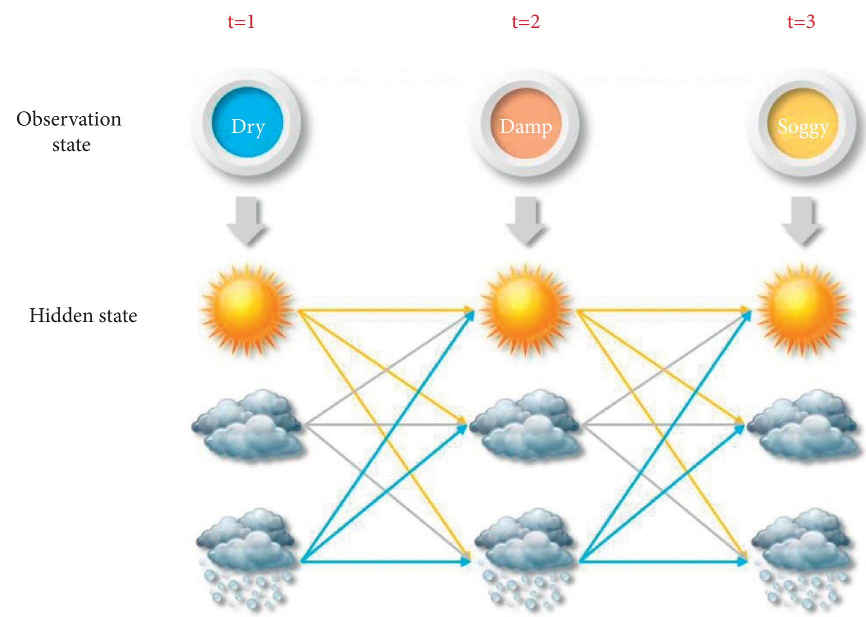


Figure 1: Hidden Markov model.

where N is the number of states. Use $q\ t=$Si to indicate that the HMM is in the hidden state Si at time t, and the hidden state sequence is

$$Q=\{q_1,\ldots,q_t\}. \quad (2)$$

(2) The probability distribution A of state transition. The probability distribution of state transition can be expressed as

$$A=\left\{P\left(q_{t+1}=S_j|q_t=S_i\right),1\le i,j\le N\right\}. \quad (3)$$

(3) Probability distribution B of the observed variable output under the condition of state Si. Assuming that the sample space of the observed variable is V, the probability distribution of the output observed variable in the state Si can be expressed as

$$B=\{f\left(Q_t=v|q_t=S_i\right),1\le i\le N,v\in V\}, \quad (4)$$

where Q t is the observed random variable at time t, which can be a value or a vector, and the observation sequence is denoted as O. It is worth noting that the sample space and probability distribution of the observed variables here can be discrete or continuous.

(4) The probability distribution of the initial state of the system π. The probability distribution of the initial state of the system can be expressed as

$$\pi=\{P\left(q=S_i\right),1\le i\le N\}. \quad (5)$$

In summary, to describe a complete HMM, the model parameters are required to be {S, A, B, π}. For simplification, it is often expressed in the following form, namely, $\lambda=$ {A, B, π}. So, figuratively speaking, HMM can be divided into two parts: one is a Markov chain, described by {π, A}, and the output is a hidden state sequence; the other random process is described by B, and the output is a sequence of observations. In addition, for a standard HMM model, three basic problems need to be solved: model training, hidden state estimation, and likelihood calculation.

(1) *Model Training*. The problem of model parameter estimation is how to adjust the parameters of the model $\lambda=$ {A, B, π} for the initial model and the given observation sequence O for training, so that it can best fit the observation data, namely:

$$\lambda_{ML}=\arg\max_{\lambda} f\left(O|\lambda\right). \quad (6)$$

This can be done by the EM algorithm or the Baum–Welch algorithm.

(2) *Likelihood Calculation*. Given the model parameter λ, calculate the likelihood of the observation sequence O, that is, calculate the likelihood f(O | λ) or log likelihood lnf (O|λ), which represents the accuracy of the parameter λ to fit the data O. This can be obtained by traversing the HMM once through the forward-backward algorithm.

(3) *Hidden State Estimation*. Given model parameters $\lambda=$ {A, B, π} and observation data O, estimate the most probable hidden state sequence Q based on a certain optimal criterion, that is, estimate the most likely path through which the observation sequence is generated. This can be done by the Viterbi algorithm. Make an estimate.

The hidden Markov model has the following three main application problems, as shown in Figure 3.

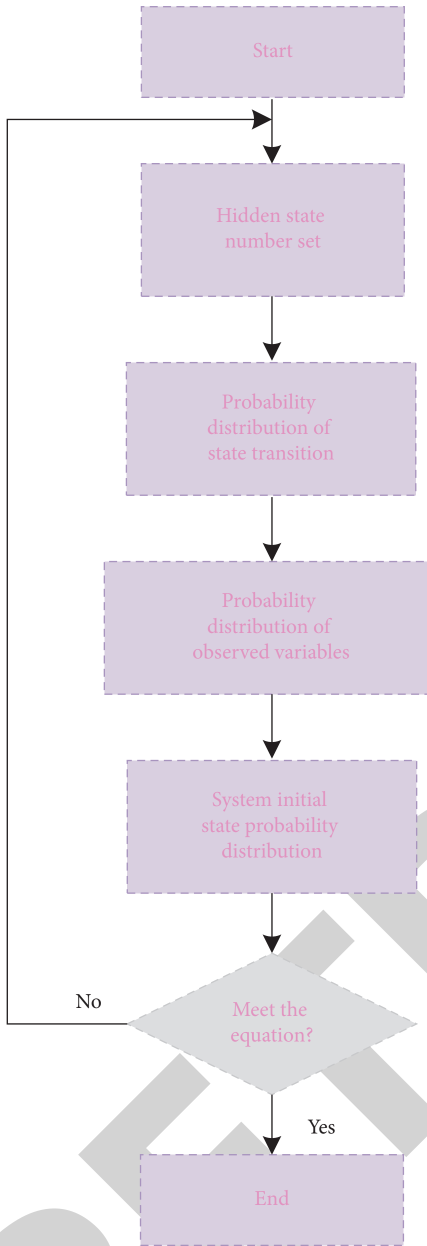


Figure 2: Flowchart of hidden Markov model.

2.1. Assessing the Problem. This type of application is based on the known hidden Markov model to find the probability of a sequence of observations.

To use forward algorithm (forward algorithm) to solve this problem, this type of problem is based on the assumption that a series of hidden Markov models are known to describe different systems, the law of weather changes in summer, and the law of weather changes in winter. We want to know which system is most likely to generate a sequence of observations. It can also be understood as applying weather systems in different seasons to a given sequence of observations, and the season corresponding to the system with the highest probability is the most likely season. This is a similar application in

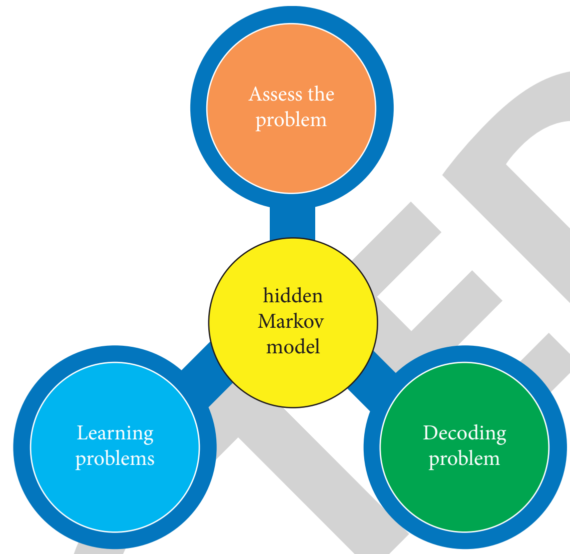


Figure 3: Three main application problems.

speech recognition systems. Define the forward variable as probability:

$$\alpha_t(i) = P(O_1, O_2, \ldots, O_t, X_t|\lambda). \tag{7}$$

Then,

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i). \tag{8}$$

It can be seen from this definition that the forward algorithm can be used to calculate the probability of a sequence of observations when the parameters of the hidden Markov model are known. In this way, in practical applications, the maximum value of P obtained by substituting the parameters of different hidden Markov models is the hidden Markov model parameter that is most consistent with the observed value sequence. The comparison of the prediction is shown in Figure 4.

2.2. Decoding Problem. This kind of application problem is to find the most likely hidden state sequence based on the sequence of observations. In many cases, people are more interested in model hidden states since they represent some more valuable things that are usually not directly observed. For example, there is a famous example of the relationship between the state of seaweed and the weather on the Internet: a blind person can only feel the state of seaweed, but he wants to know the weather. The actual state of the weather at this time is a hidden state. Therefore, the blind person needs to judge the state of the weather by observing whether the seaweed is dry or wet.

When the observation sequence and the hidden Markov model are known, the Viterbi algorithm can be used to determine or identify the most likely hidden state sequence of the sequence. A good application of the Viterbi algorithm is part-of-speech tagging in natural language processing. In

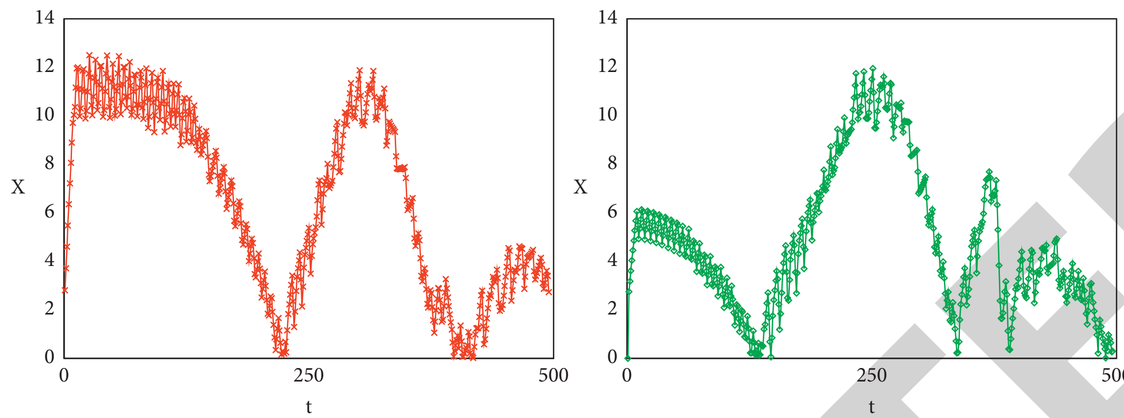


FIGURE 4: Comparison of the prediction.

the part-of-speech tagging problem, the words in the sentence are in the observed state, and the part-of-speech (grammatical category) is the hidden state (for many words, such as wind, fish has more than one part of speech). For the words in each sentence, by searching for the most likely hidden state, the most likely part-of-speech tag of each word can be found in a given context.

2.3. Learning Problems. This type of problem refers to determining the parameters of the hidden Markov model from the sequence of observations. It is a method of training a hidden Markov model using known data. This is the most commonly used tool because in practical problems, it is usually impossible to know the parameters of a ready-made hidden Markov model. These parameters include the initial matrix, the state transition matrix A, and the output matrix B. It can only be determined by studying the sequence of observations of the hidden Markov model. The evaluated data are shown in Figure 5.

This is the most difficult application problem of the hidden Markov model. A triple (Π, A, B) should be generated according to the observation sequence and the hidden state it represents, so that this triple can best describe the phenomenon we see. You can use the forward-backward algorithm to solve this problem.

The backward variable can be defined as

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \ldots, O_T, |X_t, \lambda)\lambda = (\Pi, A, B). \tag{9}$$

Let the backward variables of all states at $t = T$ be 1. Thus, the backward variable of each time point $t = T-1, T-2, \ldots,$ 1 can be calculated:

$$\beta_t(i) = \sum_{i=2}^{T} a_{ij} b_j(O_{t+1})\beta_{t+1}(j) t = T-1, T-2, \ldots 1, 1 \le i \le N. \tag{10}$$

What is calculated in this way is the backward variable corresponding to the hidden state at each moment. Given the observation sequence O and the hidden Markov model parameter λ, define the probability variable at the hidden state qi at time t as

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{n} \alpha_t(i)\beta_t(i)}. \tag{11}$$

Given the observation sequence O and the hidden Markov model, define the probability variable of being in the hidden state qi at time t and in the hidden state qj at time $t+1$ as

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_t(i) a_{ij} b_j(O_{t+1})\beta_{t+1}(j)}. \tag{12}$$

Therefore, if the above three formulas are calculated iteratively, the parameters of the hidden Markov model can be continuously re-estimated. Then, after multiple iterations, a maximum likelihood estimate of the hidden Markov model can be obtained. However, it should be noted that the result obtained by the forward-backward algorithm is a local optimal solution. The FFT analysis is shown in Figure 6.

3. Time Series

Because the causal prediction method only uses the causal relationship between a variable and other variables, it lacks the function of describing the structure of the variable's own time series, while the time series prediction method can only describe the structure of the variable's own sequence but does not consider other related factors. A sequence composed of some random variables xN is called a random sequence, which can be represented by a set {xN}. A random vector X can also be defined in a multidimensional random space, where each component is represented by xi. Then, the so-called time series are sorted in the order of time, that is to say, the subscript in xi is the variable of time t, where t is an integer, which represents the increment of the time interval, which is called a random sequence. It is also a time series. It is usually represented by {xt }. In the time variable of the time series, t can be a positive integer or a negative integer because they are all based on the current moment. A negative value means that it is generated before the current moment; a positive value means that it is generated after the current moment.

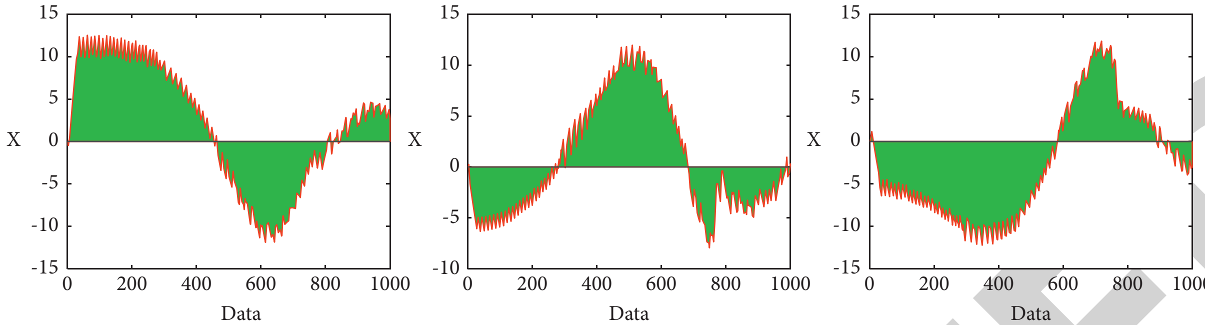


Figure 5: Evaluated data.

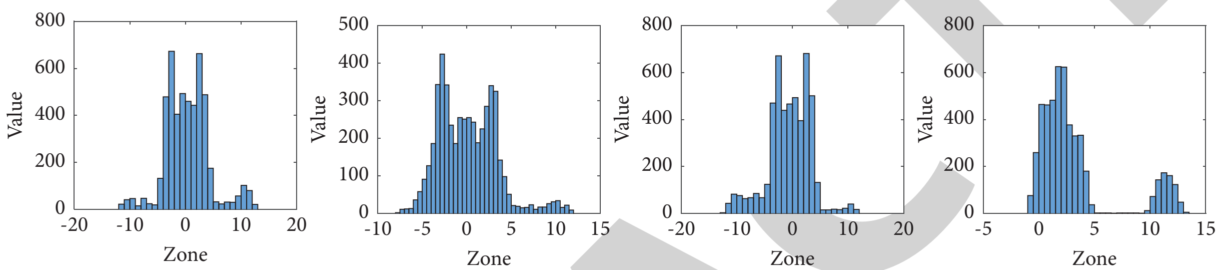


Figure 6: FFT analysis.

The content of this section is to discuss the time series model, and we use mathematical methods to express the structure of the time series model. Thus, by studying the characteristics of autocorrelation and partial autocorrelation of time series, it lays a foundation for the subsequent establishment of mathematical models, so as to better study the changing laws of volatility and ultimately enable people to predict or avoid risks in the process of financial transactions.

Usually, due to the inertia of the economic system, the time series related to the economy often have a contextual dependence. The simplest case is that the current value of the variable is mainly related to the value of the previous period. The mathematical model to describe this relationship is the first-order autoregressive model. The prediction is shown in Figure 7.

The stock market index is a complex financial time series. Behind the rise and fall of stock prices, there are many economic and noneconomic factors that hide the influence. These influences are like an invisible hand that controls the stock market. How to analyze the various influencing factors reflected behind the index and the economic status through the rise and fall of the stock market index and how to use the changes in the economic status behind the stock index to predict the stock market index are all issues worthy of study. This article will use hidden Markov model to analyze and study these two situations. By studying the stock market index to discover a certain hidden state behind the stock market, on the one hand, and by mastering and estimating the changing law of this hidden state, on the other hand, the forecasting model is improved, thereby improving the prediction effect of the model.

In this section, we will give improvements to the nonparametric kernel regression model and the least squares support vector machine model based on the hidden Markov model, that is, the mechanism conversion hybrid prediction model, and the hidden state sequence is different, and the number of samples is different. The prediction method of the mechanism conversion hybrid prediction model is expressed in three forms: LSSVM-LSSVM model, KERNEL-KERNEL model, and KERNEL-LSSVM model.

The algorithm that realizes the decoding function of the hidden Markov model is called the Viterbi algorithm. In the analysis of financial time series in this section, it is assumed that it has only two hidden states. In fact, these two hidden states can be understood as the normal state of the stock market and the state when abnormal conditions occur, just like the "barometer" of the stock market. Therefore, after this article, these two states will be referred to as normal state and abnormal state. In order to meet the research needs of this article, some improvements have been made to the previously defined Viterbi algorithm and the two-category Viterbi algorithm. Value vs. time is shown in Figure 8. As can be seen, the predicted value agrees well with the above analyses, which also means that the validation of this paper is reasonable.

Here, the alpha quantile can be achieved with the help of the QUANTILE function in the MATLAB software. This paper takes 0.8, so that the number of normal state points in the sequence is 4 times that of abnormal state points. This is more reasonable because it is clear that there should be more cases. In fact, different α can be selected in different problems.

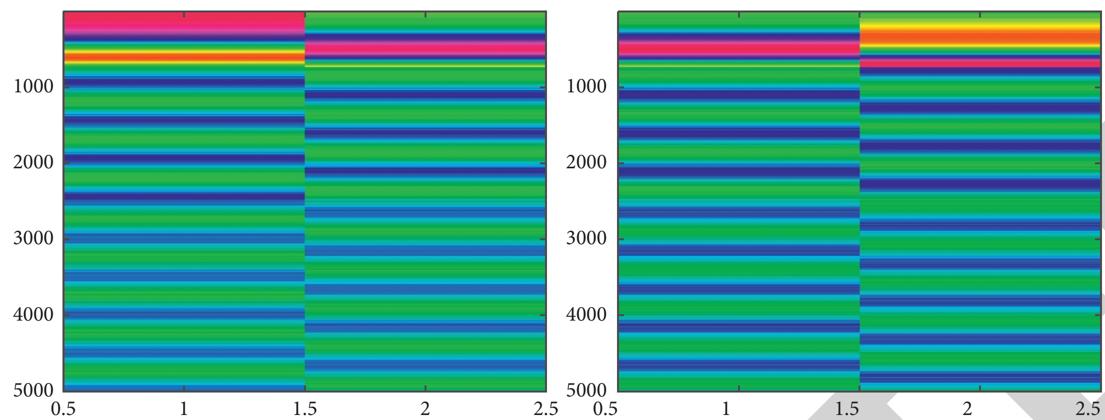


Figure 7: Prediction.

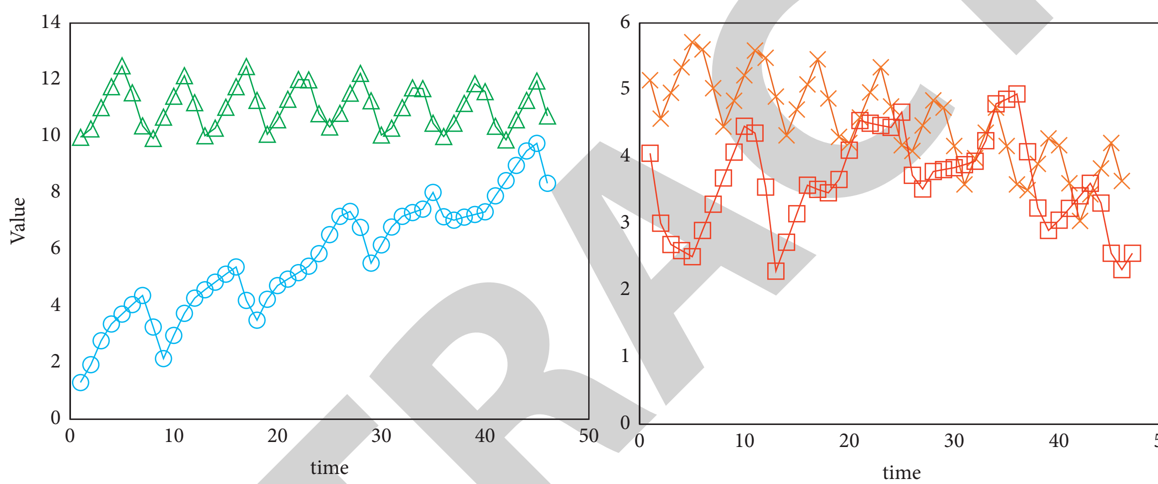


Figure 8: Value vs. time.

In addition, if you want to implement these steps in practical applications, it is obvious that you first need to use the Baum–Welch algorithm, which is the forward-backward algorithm defined in this article, to train the hidden Markov model. It is necessary to obtain the Viterbi algorithm. The parameters are the initial probability, state transition probability, and output probability. In this way, the hidden state of the sequence can be obtained by using the above Viterbi algorithm.

The previous section of this article introduced the theory of hidden Markov models. There is a kind of decoding problem in the application of hidden Markov model. This type of application is to find the most likely hidden state sequence based on the sequence of observations. For the research object of this article: financial time series, we can use this decoding function of hidden Markov model to reveal its hidden state sequence.

After using the hidden Markov model Viterbi algorithm to "decode" the hidden state of a financial time series, two hidden state series will be obtained. These are the normal state sequence and the abnormal state sequence of the time series. Since the most fundamental purpose of this article is to predict the financial time series, what must be done next is to predict the hidden state at time $T+1$ based on the hidden state before the sequence.

According to the principle of hidden Markov model, it is assumed that the hidden state sequence is a first-order Markov process. In other words, the hidden state at time $T+1$ is only related to the state at time T. Then, as long as you know the state transition probability matrix A and δ at time T, you can find δ at $T+1$. Then, compare δ when i takes different values, and note that i that maximizes δ is the state at time $T+1$.

This method can make a one-step prediction for the situation where the hidden Markov model contains two types of hidden states. In this paper, it is called the one-step hidden state prediction method of the two-class hidden Markov model. Using this algorithm, the prediction of the hidden state of the financial time series at $T+1$ is realized.

One thing to note here is that the parameters of the hidden Markov model are required to be known in the defined conditions, but this is almost impossible to achieve in practical problems. Therefore, in practical applications, we must first use the forward-backward algorithm introduced in this article to train the parameters of the hidden Markov model and then use the definitions in this section to identify and predict the hidden state of financial time series at various points in time.

After obtaining the hidden state of the sequence at time $1, \ldots, T$ and predicting the hidden state at time $T+1$, the next thing to do is to choose a suitable method to predict the financial time series. The basic methods used in this paper are still nonparametric kernel regression models and least squares support vector machine models. However, the original financial time series was decomposed into two subsequences, namely, the normal state sequence and the abnormal state sequence. For these two subsequences, different methods need to be selected for prediction according to their different properties, in order to achieve better prediction results.

This paper integrates the above-mentioned theories and models and proposes a hybrid forecasting model of mechanism transformation. Because this paper uses different methods to predict different hidden state subsequences and according to the different methods, the mechanism conversion hybrid prediction model is expressed in the following three specific forms: LSSVM-LSSVM model, KERNEL-KERNEL model, and KERNEL-LSSVM model. The difference between these three models lies in the different prediction methods for the sequence. The following are the methods for predicting the sequence of these three models according to the sample data volume of the financial time series. The predicted results are shown in Figure 9.

4. Empirical Analysis

All the empirical analyses in this article were conducted using MATLAB software. We use MATLB Hidden Markov Model Toolkit and Least Squares Support Vector Machine Toolkit: LS-SVMlab version 1.5.

This section uses the LSSVM-LSSVM and KERNEL-KERNEL model forecasting methods to conduct empirical analysis on the Shanghai Stock Exchange Index and the KERNEL-LSSVM model forecasting method to conduct empirical analysis on the NASDAQ Composite Index. Compare the results with the prediction results of ordinary LSSVM model or nonparametric kernel regression model.

The closing price data of the Shanghai Stock Exchange Index as of March 11, 2012, a total of 2932 trading days, were selected as the experimental dataset. In order to obtain better training of the support vector machine model, select the first 80% of the data as the training dataset and the last 20% of the data as the test dataset. At this time, the sample data volume is not much, so the data are predicted using the ordinary LSSVM model and the LSSVM-LSSVM model.

When using LSSVM to predict time series, it is necessary to select appropriate parameters gam and sig2 in the program. These two parameters, respectively, correspond to the constant C and variance 2 in this paper. These two parameters are determined according to the number of samples, and the value of sig2 does not have a great influence on the prediction results, so this article generally takes $\text{sig2} = 0.2$. When the ordinary LSSVM prediction is performed, gam passes the test and is determined to be 21. In LSSVM-LSSVM, 1.5 is selected when predicting the value of the normal state point. The selection of this parameter has already reflected the improvement of the method in this paper because the value of gam is affected by outliers in the sequence. The more the outliers and the greater the deviation, the larger the value of gam. Therefore, it can be determined that the hidden Markov model does separate outliers from the sequence. In addition, the LSSVM regression models used in this article are all first-order autoregressive forms.

The comparison chart of the prediction effect of the ordinary LSSVM model of the LSSVM-LSSVM model is as follows.

This improvement comes from two aspects. First, it is effective to use the hidden Markov model to decompose financial time series according to hidden states, and the decomposed normal state subsequence reduces the influence of outliers. Second, the mechanism conversion hybrid prediction model is effective in reducing the prediction error of the model.

The mechanism conversion hybrid prediction model proposed in this paper can significantly improve the prediction effects of the LSSVM model and the nonparametric kernel regression model. However, the improvement of the LSSVM model is more obvious. Moreover, it can also be seen that in the problem of predicting the Shanghai Composite Index, the prediction errors of the LSSVM model are relatively small. That is to say, it is more suitable to use LSSVM and its improved model LSSVM-LSSVM to predict the Shanghai Composite Index. In fact, in the course of the experiment, it was found that when the sample size of the financial time series is small, in the methods discussed in this article, it is always appropriate to use the LSSVM model or its improved model.

But this does not mean that the nonparametric kernel regression model has no advantages at all because whether it is the LSSVM model or the LSSVM-LSSVM model, there is a problem in the application process: the model parameters gam and sig2 are difficult to determine. In the experiment process of this article, this problem took a lot of time. As the sample size increases, the training and parameter determination time of the LSSVM model will become very long. At the same time, even if the parameter gam is found, it will be quite large, which does not meet the original intention of the least squares support vector machine. The nonparametric kernel regression model does not require much calculation, and there are no parameters that are difficult to determine. Also, it has good properties under large samples. Therefore, when the amount of financial time series data is large, the nonparametric kernel regression model can only be used to predict it.

In addition, because this paper divides the financial time series into normal state series and abnormal state series, the abnormal state series has fewer points, so the LSSVM model is considered to predict the abnormal state series. Through comparison, see if it can improve the forecasting effect. Therefore, in the next experiment, the ordinary nonparametric kernel regression model, KERNEL-KERNEL model,

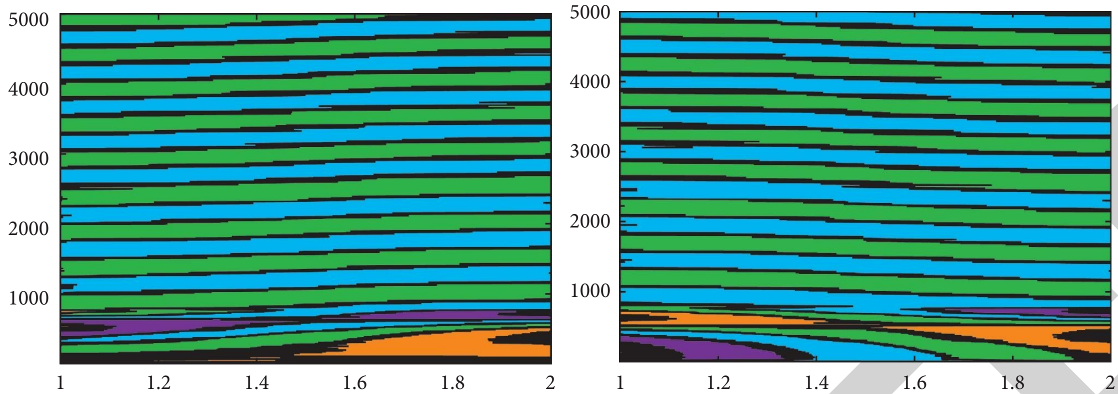


Figure 9: Predicted results.

and KERNEL-LSSVM model will be used to predict the financial time series with a large amount of data, and then their prediction effects will be compared.

5. Conclusion

This paper makes use of the assumption that the hidden state of the hidden Markov model is a first-order Markov process and proposes a one-step method to predict the next hidden state of a financial time series when the sequence has two types of hidden states. After getting the prediction result of the hidden state, this paper also proposes to choose different models to predict its value according to the different amount of financial time series sample data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] T. A. Gowan, "A hidden Markov model for estimating age-specific survival when age and size are uncertain," *Ecology*, vol. 102, no. 8, Article ID e03426, 2021.

[2] Y. Zhao, S. Shen, and H. X. Liu, "A hidden Markov model for the estimation of correlated queues in probe vehicle environments," *Transportation Research Part C*, vol. 128, 2021.

[3] Z. Wu, C. Wang, W. Peng, W. Liu, and H. Zhang, "Non-intrusive load monitoring using factorial hidden Markov model based on adaptive density peak clustering," *Energy and Buildings*, vol. 244, 2021.

[4] R. Wen, Q. Wang, and Z. Li, "Human hand movement recognition using infinite hidden Markov model based sEMG classification," *Biomedical Signal Processing and Control*, vol. 68, 2021.

[5] J. J. Jung and G. S. Jo, "Brokerage between buyer and seller agents using constraint satisfaction problem models," *Decision Support Systems*, vol. 28, no. 4, pp. 291–384, 2020.

[6] B. Alberto, L. Carbonari, P. M. Gil, and G. Trovato, "Economic growth and innovation complexity: an empirical estimation of a Hidden Markov Model," *Economic Modelling*, vol. 98, pp. 86–99, 2021.

[7] J. Joanna, M. Balcerek, K. Burnecki, A. Sabri, M. Weiss, and D. Krapf, "Identifying heterogeneous diffusion states in the cytoplasm by a hidden Markov model," *New Journal of Physics*, vol. 23, no. 5, pp. 1–9, 2021.

[8] A. Nagamani, S. Anuktha, N. Nanditha, and V. Agrawal, "A genetic algorithm-based heuristic method for test set generation in reversible circuits," *IEEE Trans. on CAD of Integrated Circuits and Systems* .vol. 37, no. 2, pp. 324–336, 2018.

[9] S. Li and J. Lian, "Hidden Markov model based control for singular Markov jump systems[J]," *Journal of the Franklin Institute*, vol. 358, no. 8, pp. 4141–4158, 2021.

[10] Y. Zhao, W. Gao, and S. Carol, "Sequential Bayesian inference of transition rates in the hidden Markov model for multi-state system degradation," *Reliability Engineering & System Safety*, vol. 1, Article ID 107662, 2021.

[11] Y. E. Shin, D. Liu, H. Sang, T. Ferguson, and P. Song, "A binary hidden Markov model on spatial network for amyotrophic lateral sclerosis disease spreading pattern analysis," *Statistics in Medicine*, vol. 40, no. 13, pp. 3035–3052, 2021.

[12] J. A. Brander and E. J. Egan, "The winner's curse in acquisitions of privately-held firms," *Review of Economics & Finance*, vol. 65, pp. 249–262, 2017.

[13] Z. Palmowski, "A note on var for the winner's curse," *Economics/Ekonomia*, vol. 15, no. 3, pp. 124–134, 2017.

[14] M. Rahul, N. Kohli, and R. Agarwal, "Facial expression recognition using local multidirectional score pattern descriptor and modified hidden Markov model," *International Journal of Advanced Intelligence Paradigms*, vol. 18, no. 4, pp. 1–8, 2021.

[15] L. Münch Jan, "Robust hidden markov model selection by bayesian network-based algorithms embracing various problems of experimental data," *Biophysical Journal*, vol. 120, no. 3S1, p. 338a, 2021.

[16] P. Klibano, M. Marinacci, and S. Mukerji, "A smooth model of decision making under ambiguity," *Econometrica*, vol. 73, no. 6, pp. 1849–1892, 2005.

[17] S. Kosuke, S. Z. Maw, T. T. Zin, P. Tin, I. Kobayashi, and Y. Horii, "Activity-integrated hidden markov model to predict calving time," *Animals*, vol. 11, no. 2, p. 385, 2021.

[18] B. P. Y. Loo, J. Hsiao, A. B. Chan, and H. Lan, "Applying the hidden markov model to analyze urban mobility patterns: an interdisciplinary approach," *Chinese Geographical Science*, vol. 31, no. 1, pp. 1–13, 2021.

[19] X. Zhou, K. Kang, T. Kwok, and X. Song, "Joint hidden markov model for longitudinal and time-to-event data with

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9764272, 1 page
https://doi.org/10.1155/2023/9764272

Retraction

Retracted: High-Order Moment Contagion of the Carbon Market: A Heterogeneity Analysis of Market Volatility Trend

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Ni, P. Yun, and J. Sun, "High-Order Moment Contagion of the Carbon Market: A Heterogeneity Analysis of Market Volatility Trend," *Security and Communication Networks*, vol. 2022, Article ID 1608871, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1608871, 14 pages
https://doi.org/10.1155/2022/1608871

Research Article

High-Order Moment Contagion of the Carbon Market: A Heterogeneity Analysis of Market Volatility Trend

Li Ni[1], **Po Yun**[2], **and Jingping Sun**[1]

[1]*School of Finance and Taxation, Anhui Finance & Trade Vocational College, Hefei 230601, Anhui, China*
[2]*School of Economics and Management, Hefei University, Hefei 230601, Anhui, China*

Correspondence should be addressed to Li Ni; nili7759121@126.com

Received 25 January 2022; Revised 3 March 2022; Accepted 10 March 2022; Published 16 April 2022

Academic Editor: Chin-Ling Chen

Information asymmetry and extreme events shocks can lead to the phenomenon of significant carbon market contagion. However, the existing studies mainly focus on the low-order moment of carbon price, making it difficult to reveal the risk contagion characteristic caused by irrational behaviors and policy shocks. This article takes market skewness and kurtosis into the research framework and constructs the FR, CS, and CK statistical model to detect the contagion in correlation channel, coskewness channel, and cokurtosis channel, respectively. The contribution of this article is to reveal the significant high-order moment contagion channel and strength of carbon market to its infected market under different market volatility trends. The results show significant contagion is widespread from the carbon market to its infected markets through the channels of coskewness and cokurtosis in different volatility trends. Additionally, the contagion strength in volatility rapid and slowly rise trend is generally higher than in the volatility rapid and slowly decline trend. That is to say, the shock of market irrationality and external events in the carbon market measured by the high-order moment contagion channels are essential risk factors that affect its infected markets. Those results convince that the acceptance of significant contagion sourced from the carbon market varies for different infected markets.

1. Introduction

As a financial tool to achieve carbon dioxide emission reduction, the carbon market was formally established in 2005, with the signing of the Kyoto protocol. The market realizes the target of resources allocation and emission reduction through market transaction among various reduction entities. The signing of the Paris Agreement in December 2015 further highlights the carbon market's capital allocation mechanism for emission reductions. The carbon market is characterized by strong sensitivity to policy shocks, and the price is vulnerable to external event shocks compared with other markets. For example, the COVID-19 caused an abnormal global economic downturn, and led to a significant drop in carbon futures and spot prices, especially as of March 23, 2020, the carbon futures prices have fallen by as much as 35% for 13 continuous trading days. Consequently, the carbon price implies more complex risk, and is prone to transmit crisis to the other closely market through the global financial network, that forms risk contagion and exhibits different contagion performance according to different market volatility. It is found that the spillover of carbon market to crude oil and natural gas market is much larger than the spillover accepted by the carbon market [1]. Therefore, it is of theoretical and practical significance to examine the risk contagion between carbon market and its infected markets and investigate the contagion of high-order moment channels characterized by market skewness and kurtosis reflected the asymmetry and extreme shocks.

The theoretical basis of this article is the asset pricing theory under market irrationality. That is, contagion is the irrational comovement after eliminating the fundamentals and rational behaviors [2]. Based on market correlation analysis, Forbes and Rigobon [3] and Renée et al. [4] put that the contagion exists if the correlation between markets increases significantly after financial shocks.

As for the research of contagion, we find previous contagion studies only focus on the low-order moment channel (market mean and variance) to examine the price information linkage and volatility spillover between carbon market and its infected markets [5] that cannot fully reveal the carbon market characteristics of peak thick tail and heterogeneity of volatility trend [6]. Actually, the aggregated excess returns can be predicted by the skewness risk premium, which is constructed to be the difference between the physical and risk-neutral skewness [7].

Therefore, the motivation of this article is focusing on the contagion caused by high-order moment attributes rather than the manner of previous low-order moment manners. That is, we take the market skewness and kurtosis reflected the asymmetry and extreme shocks into the research framework, analyze the risk contagion of high-order moment channels in the carbon market caused by irrational behaviors and external event shocks, and explore the difference of contagion under different market volatility trends and its corresponding explanations.

The remainder of this paper is organized as follows: Section 2 reviews the related literature. Section 3 introduces the methodology and data descriptions. Section 4 presents the empirical results and robustness discussions. Finally, Section 5 is the conclusion.

2. Related Literature

Although the literature related to risk contagion of carbon market is scarce, the studies on information linkage and volatility spillover between carbon market and its homogeneous market, capital market, and energy market have provided a foundation for this article.

2.1. Research on Information Spillover between Carbon Market and Its Homogeneous Market. The products of carbon market and its homogeneous market are affected by the same supply and demand. Research concluded that there is a long-term equilibrium relationship between EUA (European Union Allowance) future and CER (Certified Emission Reduction) spot in the first phase of the European Union Emissions Trading System (EU ETS); the price discovery of CER spot is more significant than that of EUA futures [8].Conducted the multivariate GARCH model, Mansanet-Bataller et al. [9] detected a significant information spillovers correlation between EUA and CER, and the correlation coefficient floats dynamically between 0.01 and 0.9. Subsequently, Chevallier [10] introduced economic recession factor into the study of spillovers relationship between EUA and CER; the results show that economic recession significantly increases the correlation between the two markets. Employed the multivariate dynamic condition correlation model (MS-DCC-GARCH), the research points a time-varying correlation and volatility spillover relationship between EUA spot and EUA futures [11], furthermore, the spillover strength is stronger than that of EUA and CER, and the EUA market has strong price discovery function compared with other markets [12].

2.2. Research on the Correlation between Carbon Market and Capital Market. The financial nature of carbon market is prominent; the properties such as negative asymmetry and positive correlation with stocks indexes during the trading session, typical of financial assets, are detected in the work of Medina and Pardo [13]. As a reflection of macroeconomics, the capital market affects the capital flow and volatility of carbon market. The carbon price is positive correlated with the stock price; the impact of carbon market on the capital market is stronger than that on the energy market [14]. Conducted the GARCH model, Reboredo [15] pointed that the crude oil prices are closely related to macroeconomy and capital markets; it is possible to transfer the capital market uncertainty to the carbon market. The exchange market also affects carbon price; employed the Copula-ARMA-GARCH model, Zhang et al. [16] depicted the risk factors correlation between carbon and exchange market. The research pointed that the potential carbon market risk is higher than that of exchange market. Researches found that there exists considerable asymmetric risk spillover between European carbon market and financial market, and the weak bidirectional causality relation between China carbon prices and energy-intensive stock indexes has also been examined [17–19].

2.3. Research on Spillover Effect between Carbon Market and Energy Market. Energy prices are the most important drivers of carbon prices due to the ability of power generators to switch between their fuel inputs [20]. According to Nazifi and Milunovich [21], the EUA futures have a causal effect on gas prices, and the electricity prices drive carbon prices. Used the GARCH model, Aatola et al. [22] argued that the short-term dynamic carbon future return is closely to natural gas return. The price of coal market has two-way causal relationship with electricity price [23]. Hammoudeh et al. [14] described that the impact of coal price on carbon market price in falling trend is stronger than that of rising trend. Furthermore, conducted the EMD decomposition model, Cao et al. [24] hold that the relationship between carbon price and electricity price has changed from bidirectional linear causality to one-way causality with nonlinear characteristics. It is found that there exists an obvious positive relationship between the EUA and oil markets and such dynamic spillover effect varies with time [25]. Employed the canonical vine copula model, Uddin et al. [26] investigated a stronger one-way and two-way risk spillover relationship between carbon market and energy market. The time-varying and directional spillover between carbon and energy markets have detected the electricity market is the main net information receiver affected by the carbon market [27]. Additionally, the complex time-frequency and neural network mechanism between carbon and oil markets has been explored by the model of novelty partial wavelet and deep learning models [28–30].

2.4. The Comment on the Previous Studies. The common perspective of above literature, however, is based on the view of low-order moment attribute (mean-variance) of carbon asset, focusing on studying the information transmission and risk contagion effect between carbon market and its infected markets. The foundation of these studies may ignore the impact of high-order moment attribute (market skewness and kurtosis) of carbon price on the contagion process; it is difficult to capture and explain the contagion behavior of carbon market caused by external policy and irrational investment.

To overcome these shortcomings, Renée et al. [31] and Chan et al. [32] introduce the coskewness, covolatility, and cokurtosis statistics into the high-order moment CAPM framework; the contribution of the above studies is constructing a new nonparametric high-order moment contagion statistics with the assumption of chi-square distribution to measure the risk contagion. However, the research of Renée et al. [31] only consider the contagion caused by the financial crisis and other political events that may have ignored the contagion pattern caused by different market volatility states and trends. As we know, the rapid volatility of carbon market may lead to greater risk contagion compared with the slow volatility trend. Therefore, it is very necessary to provide new convinced evidence of contagion from the high-order moment perspective.

The innovation of this paper is as follows: firstly, this paper investigates the risk contagion sourced from the carbon market from the perspective of high-order moment contagion channels, which is different from the manner of low-order moment. Secondly, we judge the statistical significance of high-order moment contagion channel before investigating the contagion strength. Thirdly, we explore the structural differences of risk contagion under different market volatility trends and its corresponding explanations.

The research is based on the following steps for achieving the research objectives: firstly, the market volatility is divided into three states: stable, high-volatility and low-volatility. Secondly, the volatility further divides into rapid volatility trend (rapid rise and rapid decline) and slow volatility trend (slow rise and slow decline) and thirdly investigates the risk contagion from high-order moment channels under different market volatility trends. The logic frame of this paper is shown in Figure 1.

3. Methodology and Data Descriptions

3.1. Dividing the Volatility State. The volatility state of carbon market is uncertain; the transformation between different states is unobserved according to the complexity of carbon market volatility. Therefore, this paper extends the states of Markov State Transition model to M regimes and establishes the MS(M)-AR(P) model, which overcome the defect of same variance assumption. The model of MS(M)-AR(P) is expressed as follows:

$$R_t = \nu(M_t) + \sum_{i=1}^{p} \phi_a(M_t) R_{t-a} + \varepsilon_t, \tag{1}$$

where R_t denotes the carbon market return; ε_t represents the model residual, $\varepsilon_t \sim N(0, \sigma(M_t)^2)$; t describes the state number of carbon market volatility, $t \in \{1, 2, \ldots, k\}$; M_t obeys the first-order Markov chain; the conversion probability of M_t is expressed as follows: $P_{\text{ab}} = \text{pr}(M_t = b \mid M_{t-1} = a, M_{t-2} = \alpha, \ldots) = \text{pr}(M_t = b \mid M_{t-1} = a)$; a and b represents two random state variables and $\{a, b\} \in t$; $\nu(M_t)$, $\phi_a(M_t)$, and $\sigma(M_t)$ represents the intercept term, autoregressive coefficient, and standard deviation of carbon returns under the state of M_t, respectively.

Under the assumption of normal distribution of residual series, the conditional probability density of return R_t in regime M_t is as follows:

$$f(R_t \mid M_t = b, I_{t-1}; \theta) = \frac{1}{\sqrt{2\pi}\sigma(b)} \exp\left[\frac{-(R_t - \nu(b))^2}{2\sigma^2(b)}\right]. \tag{2}$$

When the probability of $f(M_t = b|I_{t-1}; \theta)$ is known, the probability density of R_t under the complete information condition I_{t-1} is expressed as follows:

$$\begin{aligned} f(R_t \mid I_{t-1}; \theta) = {} & p(M_t = 1 \mid I_{t-1}; \theta) f(R_t \mid M_t = 1, I_{t-1}; \theta) \\ & + p(M_t = 2 \mid I_{t-1}; \theta) f(R_t \mid M_t = 2, I_{t-1}; \theta) \\ & + \cdots + p(M_t = k \mid I_{t-1}; \theta) f(R_t \mid M_t = k, I_{t-1}; \theta), \end{aligned} \tag{3}$$

where I_{t-1} represents the value of all variables R_t in state M_t up to time $t-1$, that is, all the information contents that can be obtained up to time $t-1$, and $\theta = \{p_{\text{ab}}, \nu_i(M_t), \phi_a(M_t), \sigma_a(M_t)\}$ represents the set of model parameters, which can be estimated by the logarithmic likelihood function of the MS(M)-AR(P) model.

To identify the return series of carbon market corresponding to the certain state presented by the maximum smoothing probability, this paper takes 0.5 as the critical value of smoothing probability of each state. The sample selection is based on $p(M_t = b \mid I_T; \theta) > 0.5 \Rightarrow R(\text{state}_t)$.

3.2. Designing the Risk Contagion Model. According to the connotation of risk contagion proposed by Andrew [2], Forbes and Rigobon [3] and Renée et al. [4] put a significant method for detecting contagion; that is, the contagion exists if the correlation between markets increases significantly after financial shocks.

Therefore, the methodology of this paper adopts this idea of risk detection and conducts the statistics model of high-order moment contagion proposed by Zhang et al. [29] to solve the problem of risk contagion in the carbon market. That is, we employed the FR, CS, and CK statistical model to detect the contagion in correlation channel, coskewness channel, and cokurtosis channel, respectively. The difference between this paper and Renée et al. [31] is to detect the risk contagion caused by market volatility heterogeneity, rather than just considering the financial crisis.

3.2.1. Risk Contagion Model of the Correlation Channel. The risk contagion model of correlation channel measures the low-order moment correlation of return between carbon market and its infected markets; the model tests for a significant increase in this correlation coefficient after a shock. The cross-market correlation coefficients show as follows:

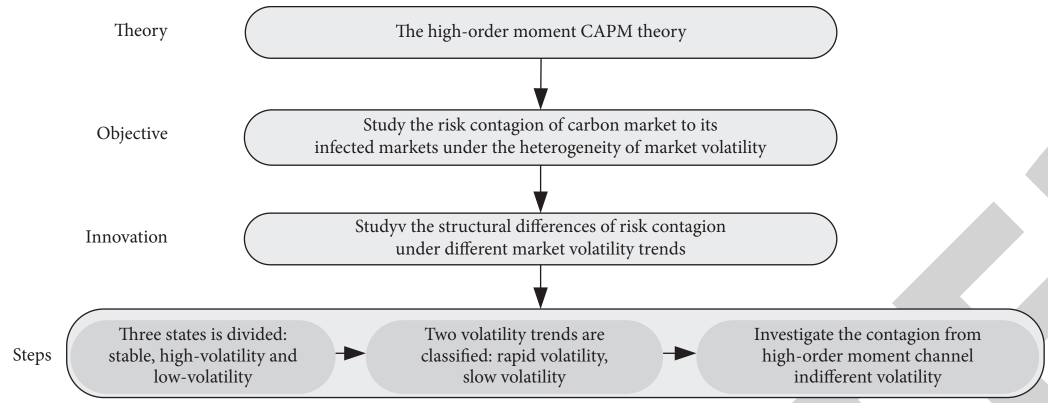


FIGURE 1: The logic frame of this article.

$$\mathrm{FR}(i \longrightarrow j) = \left(\frac{\widehat{v}_{y/x_i} - \widehat{\rho}_x}{\sqrt{\mathrm{Var}\left(\widehat{v}_{y/x_i} - \widehat{\rho}_x\right)}}\right)^2,$$
$$\widehat{v}_{y/x_i} = \frac{\widehat{\rho}_y}{\sqrt{1 + \left(s_{y,i}^2 - s_{x,i}^2/s_{x,i}^2\right)\left(1 - \widehat{\rho}_y^2\right)}}, \tag{4}$$

where i and j represents the sourced market (carbon market) and infected market (carbon homogeneous market, capital market, and energy market); x and y represents the volatility state of carbon market; and $\widehat{v}_{y/x_i}$ indicates the market correlation coefficient after the transition of volatility state. $\widehat{\rho}_x$ and $\widehat{\rho}_y$ denotes the unconditional correlation coefficient of the two markets under different market volatility while $s_{x,i}^2$ and $s_{y,i}^2$ denotes the variance of the sourced markets. If the correlation coefficient increases significantly, this suggests that the transmission mechanism between the two markets strengthened after the shock and contagion occurred.

To test a significant change in correlation coefficients of *FR* between carbon market and its infected markets under different volatility trends, the null and alternative hypotheses of *FR* are as follows:

$$H(\mathrm{FR})_0\colon \widehat{v}_{y/x_i} = \widehat{\rho}_x,$$
$$H(\mathrm{FR})_1\colon \widehat{v}_{y/x_i} \neq \widehat{\rho}_x. \tag{5}$$

Under the null hypothesis of no contagion, tests of contagion based on changes in the channel of FR are asymptotically distributed as follows:

$$\mathrm{FR}(i \longrightarrow j) \xrightarrow{\mathrm{df}} \chi_1^2. \tag{6}$$

3.2.2. Risk Contagion Model of the Coskewness Channel. The coskewness contagion model is a measure of whether the asymmetry of portfolio return between carbon market and its infected markets has changed significantly during different market volatility trend [31]. According to the difference between the market horizontal return and square return in calculating the coskewness coefficient, the test of coskewness is divided into two categories: CS_{12} and CS_{21}, where CS_{12} represents the transmission from the return of carbon market to the variance of infected markets and CS_{21} represents the transmission of carbon market variance to the return of infected markets. The smaller contagion coefficient indicates the joint distribution of portfolio is close to the standard distribution and faces less asymmetric risk.

$$\mathrm{CS}_{12}\left(i \longrightarrow j; r_i^1, r_j^2\right) = \left(\frac{\widehat{\psi}_y\left(r_i^1, r_j^2\right) - \widehat{\psi}_x\left(r_i^1, r_j^2\right)}{\sqrt{\left(4\widehat{v}_{y/x_i}^2 + 2\right)/T_y + \left(4\widehat{\rho}_x^2 + 2\right)/T_x}}\right)^2,$$
$$\mathrm{CS}_{21}\left(i \longrightarrow j; r_i^2, r_j^1\right) = \left(\frac{\widehat{\psi}_y\left(r_i^2, r_j^1\right) - \widehat{\psi}_x\left(r_i^2, r_j^1\right)}{\sqrt{\left(4\widehat{v}_{y/x_i}^2 + 2\right)/T_y + \left(4\widehat{\rho}_x^2 + 2\right)/T_x}}\right)^2, \tag{7}$$

where

$$\widehat{\psi}_y\left(r_i^1, r_j^2\right) = \frac{1}{T_y}\sum_{t=1}^{T_y}\left(\frac{y_{i,t} - \widehat{\mu}_{\mathrm{yi}}}{\widehat{\sigma}_{\mathrm{yi}}}\right)^1\left(\frac{y_{j,t} - \widehat{\mu}_{\mathrm{yj}}}{\widehat{\sigma}_{\mathrm{yi}}}\right)^2,$$
$$\widehat{\psi}_y\left(r_i^2, r_j^1\right) = \frac{1}{T_y}\sum_{t=1}^{T_y}\left(\frac{y_{i,t} - \widehat{\mu}_{\mathrm{yi}}}{\widehat{\sigma}_{\mathrm{yi}}}\right)^2\left(\frac{y_{j,t} - \widehat{\mu}_{\mathrm{yj}}}{\widehat{\sigma}_{\mathrm{yi}}}\right)^1,$$
$$\widehat{\psi}_x\left(r_i^1, r_j^2\right) = \frac{1}{T_x}\sum_{t=1}^{T_x}\left(\frac{x_{i,t} - \widehat{\mu}_{\mathrm{xi}}}{\widehat{\sigma}_{\mathrm{xi}}}\right)^1\left(\frac{x_{j,t} - \widehat{\mu}_{\mathrm{xj}}}{\widehat{\sigma}_{\mathrm{xi}}}\right)^2,$$
$$\widehat{\psi}_x\left(r_i^2, r_j^1\right) = \frac{1}{T_x}\sum_{t=1}^{T_x}\left(\frac{x_{i,t} - \widehat{\mu}_{\mathrm{xi}}}{\widehat{\sigma}_{\mathrm{xi}}}\right)^2\left(\frac{x_{j,t} - \widehat{\mu}_{\mathrm{xj}}}{\widehat{\sigma}_{\mathrm{xi}}}\right)^1. \tag{8}$$

In the above model, $\widehat{\psi}_x$ and $\widehat{\psi}_y$ represents the market skewness correlation coefficient between market i and j in volatility state x and y. r_i^1 and r_j^2 denotes the first and second order moment of market i and j, while r_i^2 and r_j^1 denotes the second and first order moment of market i and j, respectively. T_x and T_y means the market capacity under different volatility; $x_{i,t}$, $x_{j,t}$, $y_{i,t}$, and $y_{j,t}$ represents the return of sourced market and infected market under the state of x and y, respectively; $\widehat{\mu}_{\mathrm{xi}}$, $\widehat{\mu}_{\mathrm{xj}}$, $\widehat{\mu}_{\mathrm{yi}}$, and $\widehat{\mu}_{\mathrm{yj}}$ represents the mean corresponding to the above returns, respectively; $\widehat{\sigma}_{\mathrm{xi}}$, $\widehat{\sigma}_{\mathrm{xj}}$, $\widehat{\sigma}_{\mathrm{yi}}$, and $\widehat{\sigma}_{\mathrm{yj}}$ means the standard deviation of the above returns, respectively; $\widehat{v}_{y/x_i}$ indicates the market correlation coefficient after the transition of volatility state.

To test for a significant change in coskewness between carbon market and its infected markets under different volatility trends, the null and alternative hypotheses of CS_{12} are as follows:

$$\begin{aligned} H(CS_{12})_0&: \hat{\psi}_y(r_i^1, r_j^2) = \hat{\psi}_x(r_i^1, r_j^2),\\ H(CS_{12})_1&: \hat{\psi}_y(r_i^1, r_j^2) \neq \hat{\psi}_x(r_i^1, r_j^2). \end{aligned} \tag{9}$$

The null and alternative hypotheses of CS_{21} are as follows:

$$\begin{aligned} H(CS_{21})_0&: \hat{\psi}_y(r_i^2, r_j^1) = \hat{\psi}_x(r_i^2, r_j^1),\\ H(CS_{21})_1&: \hat{\psi}_y(r_i^2, r_j^1) \neq \hat{\psi}_x(r_i^2, r_j^1). \end{aligned} \tag{10}$$

Under the null hypothesis of no contagion, tests of contagion based on changes in coskewness are asymptotically distributed as follows:

$$CS_{12}(i \longrightarrow j) \xrightarrow{df} \chi_1^2, CS_{21}(i \longrightarrow j) \xrightarrow{df} \chi_1^2. \tag{11}$$

3.2.3. Risk Contagion Model of the Cokurtosis Channel. The cokurtosis contagion model is a measure of whether the portfolio between carbon market and its infected markets is affected by policy shocks or external events. Similar to the coskewness contagion model, this study divides the cokurtosis contagion test into two categories: CK_{13} and CK_{31}, where CK_{13} represents the contagion of carbon return to the skewness of infected markets, and CK_{31} means the contagion of carbon market skewness to infected market return. The higher contagion coefficient indicates that portfolio returns face greater impact of systematic risk, while the smaller coefficient indicates lower systemic risk.

$$\begin{aligned} CK_{13}(i \longrightarrow j; r_i^1, r_j^3) &= \left(\frac{\hat{\xi}_y(r_i^1, r_j^3) - \hat{\xi}_x(r_i^1, r_j^3)}{\sqrt{(18\hat{v}^2_{y/x_i} + 6)/T_y + (18\hat{\rho}_x^2 + 2)/T_x}}\right)^2,\\ CK_{31}(i \longrightarrow j; r_i^3, r_j^1) &= \left(\frac{\hat{\xi}_y(r_i^3, r_j^1) - \hat{\xi}_x(r_i^3, r_j^1)}{\sqrt{(18\hat{v}^2_{y/x_i} + 6)/T_y + (18\hat{\rho}_x^2 + 2)/T_x}}\right)^2, \end{aligned} \tag{12}$$

where

$$\begin{aligned} \hat{\xi}_y(r_i^1, r_j^3) &= \frac{1}{T_y}\sum_{t=1}^{T_y}\left(\frac{y_{i,t} - \hat{\mu}_{yi}}{\hat{\sigma}_{yi}}\right)^1\left(\frac{y_{j,t} - \hat{\mu}_{yj}}{\hat{\sigma}_{yi}}\right)^3 - (3\hat{v}_{y/x_i}),\\ \hat{\xi}_y(r_i^3, r_j^1) &= \frac{1}{T_y}\sum_{t=1}^{T_y}\left(\frac{y_{i,t} - \hat{\mu}_{yi}}{\hat{\sigma}_{yi}}\right)^3\left(\frac{y_{j,t} - \hat{\mu}_{yj}}{\hat{\sigma}_{yi}}\right)^1 - (3\hat{v}_{y/x_i}),\\ \hat{\xi}_x(r_i^1, r_j^3) &= \frac{1}{T_x}\sum_{t=1}^{T_x}\left(\frac{x_{i,t} - \hat{\mu}_{xi}}{\hat{\sigma}_{xi}}\right)^1\left(\frac{x_{j,t} - \hat{\mu}_{xj}}{\hat{\sigma}_{xi}}\right)^3 - (3\hat{\rho}_x),\\ \hat{\xi}_x(r_i^3, r_j^1) &= \frac{1}{T_x}\sum_{t=1}^{T_x}\left(\frac{x_{i,t} - \hat{\mu}_{xi}}{\hat{\sigma}_{xi}}\right)^3\left(\frac{x_{j,t} - \hat{\mu}_{xj}}{\hat{\sigma}_{xi}}\right)^1 - (3\hat{\rho}_x). \end{aligned} \tag{13}$$

In the above model, $\hat{\xi}_x$ and $\hat{\xi}_y$ represents the market kurtosis coefficient between market *i* and *j* in volatility state *x* and *y*. r_i^1 and r_j^3 denotes the first and third order moment of market *i* and *j* while r_i^3 and r_j^1 denotes the third and first order moment of market *i* and *j*, respectively. Other definition of parameters is consistent with the coskewness contagion model defined above. To test the risk contagion of cokurtosis channel between carbon market and its infected markets under different volatility trends, the null and alternative hypotheses of CK13 are as follows:

$$\begin{aligned} H(CK_{13})_0&: \hat{\xi}_y(r_i^1, r_j^3) = \hat{\xi}_x(r_i^1, r_j^3),\\ H(CK_{13})_1&: \hat{\xi}_y(r_i^1, r_j^3) \neq \hat{\xi}_x(r_i^1, r_j^3). \end{aligned} \tag{14}$$

The null and alternative hypotheses of CK_{31} are as follows:

$$\begin{aligned} H(CK_{31})_0&: \hat{\xi}_y(r_i^3, r_j^1) = \hat{\xi}_x(r_i^3, r_j^1),\\ H(CK_{31})_1&: \hat{\xi}_y(r_i^3, r_j^1) \neq \hat{\xi}_x(r_i^3, r_j^1). \end{aligned} \tag{15}$$

Under the null hypothesis of no contagion, tests of contagion based on changes in cokurtosis are asymptotically distributed as follows:

$$CK_{13}(i \longrightarrow j) \xrightarrow{df} \chi_1^2, CK_{31}(i \longrightarrow j) \xrightarrow{df} \chi_1^2. \tag{16}$$

3.3. Samples and Data Preprocessing. The carbon market is not only closely related to carbon homogeneous market and capital market but also related to the energy market [33]. Furthermore, this article chooses the EUAs (European Union Allowance spot) as the representation of the carbon homogeneous market. The others infected market of capital market and energy market are shown in Table 1. We choose the traded products of coal, oil, natural gas, and electricity markets as the energy market variables, and the data sourced from the Wind Database. The research samples with the period from June 2, 2009, to March 23, 2020, and there are total of 2768 samples by

Table 1: The designing of research samples.

Specific market	Typical market products	Abbreviation	Meaning of market products
Panel A: sourced market and data from intercontinental exchange (ICE)			
Carbon market	EUA future market	EUAf	Settlement price of continuous futures contract of EUA
Panel B: infected market and data from Wind Database			
Carbon homogeneous market	EUA spot market	EUAs	Settlement price of continuous spots contract of EUA
Capital market	Dow jones industrial average	DJIA	Closing price of DJIA from the US
	EURUSD	EURUSD	Closing price of EURUSD
	USD-index	USDX	Closing price of USDX
Energy market	Coal market	Coal	Futures settlement price of British thermal coal
	Brent oil market	Oil	Futures settlement price of Brent crude oil
	Natural gas market	Gas	UK natural gas continuous futures prices
	Electricity market	Electricity	US electricity retail prices

eliminating the sample missing and time inconsistency. The return is expressed as R_t, and $R_t = 100\times(\ln P_t\text{-}\ln P_{t-1})$, where P_t denotes the price of market products.

This article defines the contagion of high-order moment channel in the carbon market as a change caused by irrational behavior and policy shocks; this definition makes the sample's return contain more transaction noise, for example, the trading psychology and behavior information. To solve this problem, this research uses the VAR model to fit the original return series according to the suggestion of Forbes and Rigobon [3] and takes the residual as the substitution for measuring the contagion to control the influence of market fundamentals; the improved model of VAR is expressed as follows:

$$Z_t = \omega(L)Z_t + \varepsilon_t, \tag{17}$$

where $Z_t = \left\{R_{\mathrm{it}}, R_{\mathrm{jt}}\right\}$ is the combined return sequence of the source and infected market during the state transformation; ε_t is the residual sequence for computing the contagion statistics; $\omega(L)$ is a vector of lags; the VAR lag order is determined according to the Akaike information criterion (AIC) and Bayesian information criterion (BIC).

4. Empirical Results and Discussion

The empirical test of methodology is carried out (as showed in Figure 2) according to the following steps:

Step 1: dividing the volatility state of carbon market into stable volatility (S1), high volatility (S2), and low volatility (S3) and then clarifying the volatility trend into the rapid volatility trend (S1–S2) and slow volatility trend (S1–S3)

Step 2: examining whether there is high-order moment contagion relationship between carbon market and its infected market from the channels FR, CS, and CK, respectively

Step 3: summarizing the contagion direction and strength in the carbon market and testing the robustness of above conclusion

4.1. Identifying the Volatility Trend of Carbon Market

4.1.1. Selecting the State Transformation Model. To divide the market state accord with the characteristics of carbon price volatility and avoid the errors caused by setting the state parameter in subjectively, the performance of alternative models under different volatility states is compared according to the AIC and BIC minimization principle. As shown in Table 2, it is found that the model of MS (3)-AR (3) is more suitable for the state division of the carbon market than other models.

4.1.2. Clarifying the Volatility State and Trend. Research result shown in Table 3 reveals that the standard deviations of the three states are 1.17%, 6.94%, and 2.39%, respectively, which can define as the state of market stability, high volatility, and low volatility according to the estimation results of MS (3)-AR (3) model. Furthermore, the standard deviation of high volatility is equal to three times of the low volatility state and six times of stable volatility state, and the value of low volatility state is two times of stable volatility state. As a result, the volatility in different states of carbon market varies greatly and makes the impact of systemic risk on market returns vary fiercely.

Based on the volatility states divided above, the difference of volatility coefficient between state1 and state 2 is 5.77%, 1.22% for S1 and S3, and 4.55% for S2 and S3. Therefore, the state transformation difference of 5.77% and 1.22% is highly representative in denoting the maximum and minimum market volatility state differences. Based on this, we consider the volatility difference of the three states; this article defines the transformation between S1 and S2 and S1 and S3 as the representative transmission channel for measuring the risk contagion. Correspondingly, those two channels further divide into two trends: rapid volatility and slow volatility. Among them, the rapid change of volatility is divided into two kinds of trends: volatility rise rapidly (contagion from stable state to high volatility state, S1–S2) and volatility decline rapidly (contagion from high volatility state to stable state, S2–S1); and slow volatility is divided into volatility rise slowly (contagion from stable state to slow volatility state, S1–S3) and volatility decline slowly (contagion from slow volatility state to stable state, S3–S1).

Figures 3–5, respectively, show the relationship between carbon assets and standard price changes of the homogeneous market, capital market, and energy market. It can be found that the market price trend of carbon assets and homogeneous products is basically the same regardless of

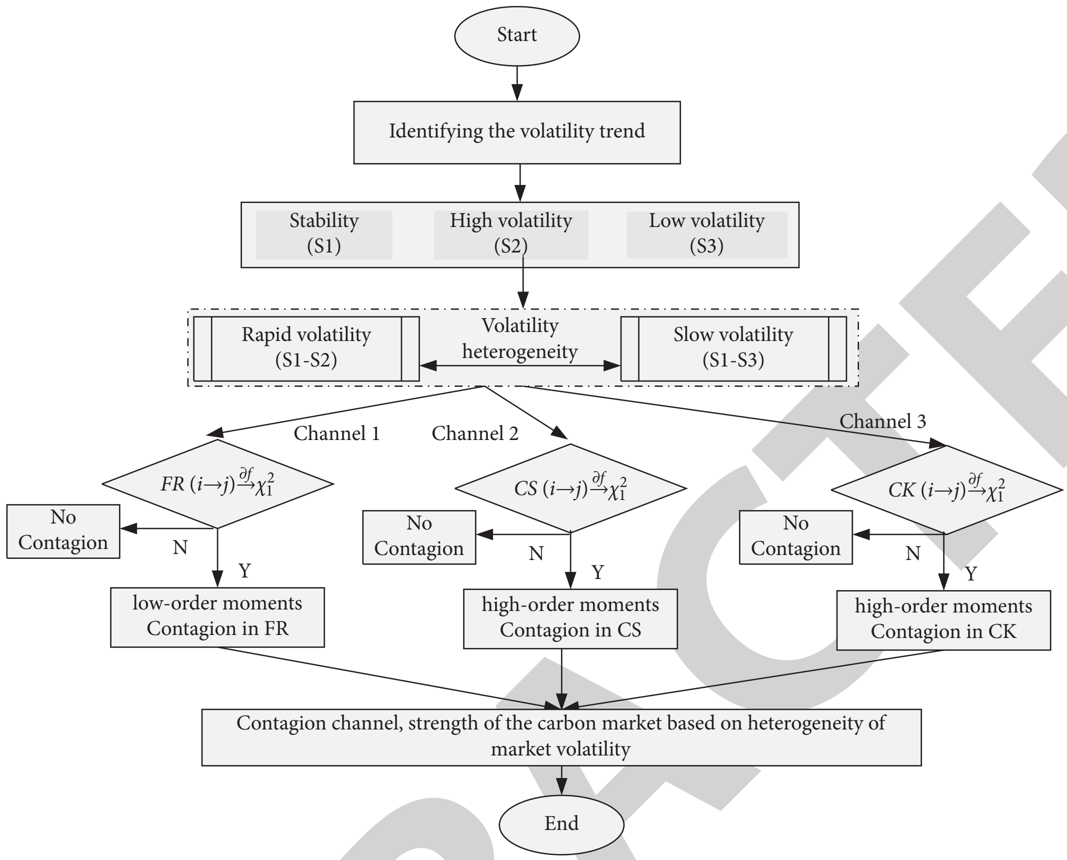


Figure 2: The flowchart of empirical design in this paper.

Table 2: Performance comparison of different state transition models of the carbon market.

Alternative model	Residual distribution	Number of parameters	Likelihood value	AIC	BIC
MS(2)-AR(3)	T	16	6477.4992	−12922.9984	−12899.9263
	N	14	6379.8549	−12731.7098	−12711.5217
MS(2)-AR(4)	T	18	6375.5774	−12715.1548	−12689.1986
	N	16	6380.1796	−12728.3592	−12705.2871
MS(3)-AR(3)	T	27	6487.9839	−12921.9678	−12883.0336
	N	24	6517.1826	−12986.3652	−12951.7569
MS(3)-AR(4)	T	30	6488.3855	−12916.771	−12873.5107
	N	27	6517.7822	−12981.5644	−12942.6302
MS(4)-AR(3)	T	40	6518.2154	−12956.4308	−12898.7504
	N	36	6465.1356	−12858.2712	−12806.3589
MS(4)-AR(4)	T	44	6510.2059	−12932.4118	−12868.9634
	N	40	6526.3629	−12972.7258	−12915.0454

the change of volatility state. This result is basically consistent with the research of Chevallier [10] that the EUA future price guides the spot price. The possible reason is that carbon assets and their homogeneous products have the same trading attributes, their price driving mechanism is relatively similar, and futures products have a strong price discovery function for spot products [27].

4.2. Analyzing High-Order Moments Statistics. Research result shows in Table 4, in term of the portfolio with carbon homogeneous products, as the volatility changes from stable to high volatility (S1–S3–S2), the increasing risk lead the coskewness coefficient of EUAf and EUAs decrease gradually and turn to negative; the result indicates that the return of portfolio composed of EUA futures and spot has significant asymmetry effect; that is, the probability of return decline is greater than that of rise. Additionally, the increasing of cokurtosis coefficient means the return faces increasing external risk shocks.

As for the portfolio with the capital market, Table 4 shows negative coskewness statistic and decreased significantly as the increasing of market volatility. The result proves that the portfolio return has significant asymmetry; the

Table 3: Parameters estimation of state transition of the carbon market based on MS (3)-AR (3) model.

State	Coefficient of volatility (%)	State description	Transition probabilities	Duration	Standard error	*P* value
S1	1.17***	Stability	0.99	74.98	0.0003	≤0.000
S2	6.94***	High volatility	0.88	8.11	0.0029	≤0.000
S3	2.39***	Low volatility	0.97	39.01	0.0004	≤0.000

Note. ***The statistical significance at the 1% level.

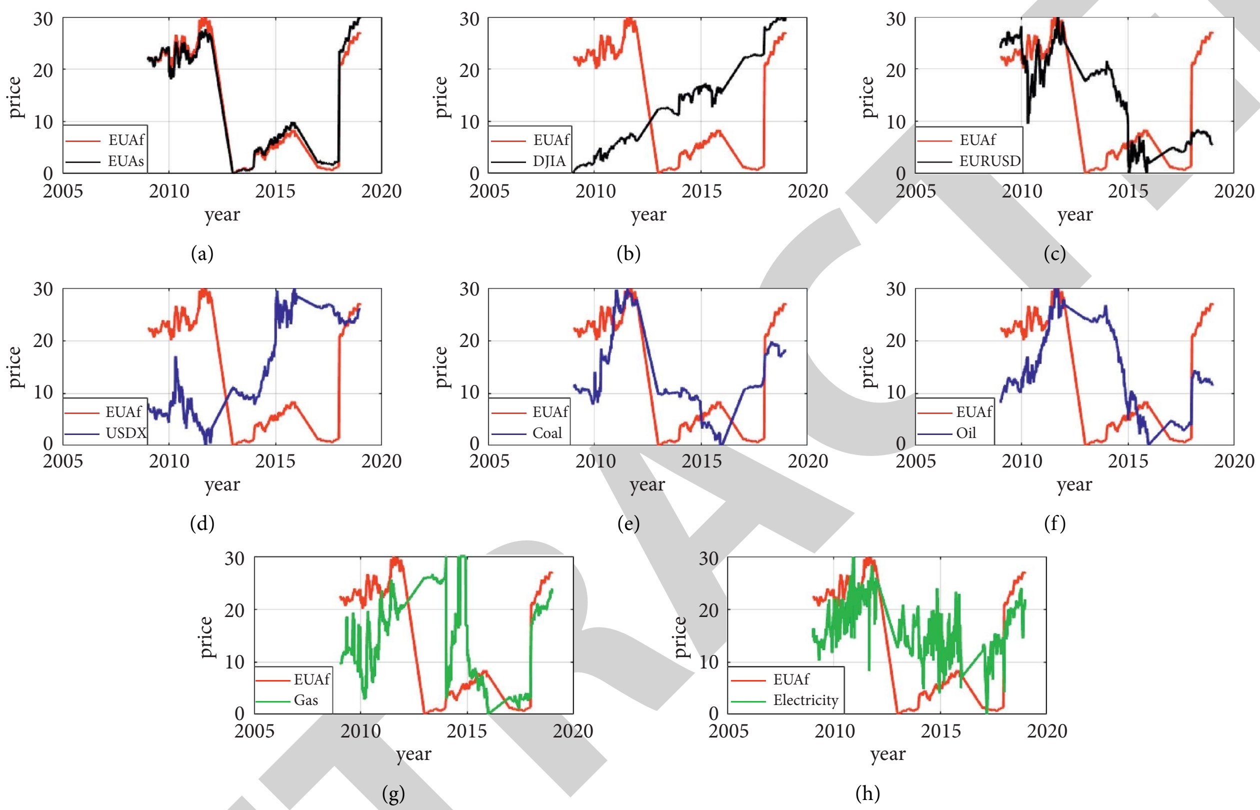


Figure 3: Standardized price change between carbon asset and its pricing factors in stable state (S1). (a) EUAf-EUAs. (b) EUAf-DJIA. (c) EUAf-EURUSD. (d) EUAf-USDX. (e) EUAf-coal. (f) EUAf-oil. (g) EUAf-gas. (h) EUAf-electricity.

probability of return decline and rise is quite different according to the coskewness coefficient. For the coefficient of cokurtosis, however, the portfolio suffers less impact from the external events because the cokurtosis is small in most cases. Therefore, the volatility of the carbon market and capital market is basically stable; the portfolio of those two markets can avoid investment risks effectively.

As for the portfolio with the energy market, the cokurtosis of joint distribution between EUAf and gas is the largest and negative in the stable stage, the result indicates that the distribution of those two markets is extremely different, and they are not effective substitutions for portfolio.

4.3. Analyzing the Risk Contagion of the High-Order Moment Channel. We find a valuable conclusion from Tables 5 and 6 that there is a significant risk contagion effect from the carbon market to its infected markets in majority of the high-order moment channels rather than the FR contagion in low-order perspective both in rapid and slow volatility trend. This conclusion convinced that the shock of market irrationality and external events measured by the high-order moment contagion channels are essential risk factors that affect its infected markets. An extended finding is that we may get wrong or inaccurate conclusions if only the correlation coefficient in the view of low-order moments is used to judge the existence of risk contagion.

4.3.1. Analyzing the Risk Contagion in the Rapid Volatility Trend. Empirical results showed that, as in Table 5, when the market in trend of rapid volatility, the number of significance in the high-order moment channels is the largest compared with other infected markets as for the homogeneous markets. There is a significant risk contagion from the carbon market to its homogeneous markets in all the high-order moment channels along with the significant FR contagion in the low-order perspective. This conclusion proved that the shock of market irrationality and external

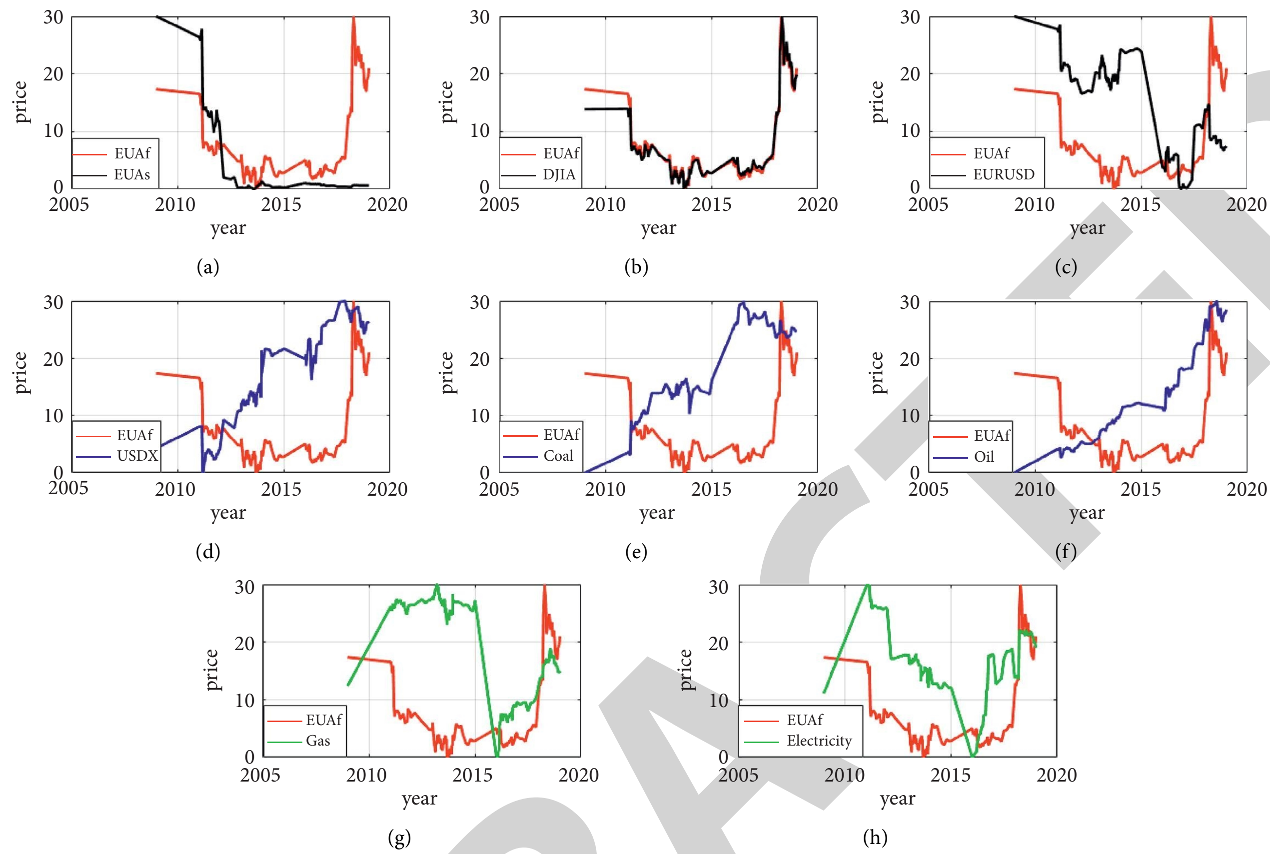


Figure 4: Standardized price change between carbon asset and its pricing factors in high volatility state (S2). (a) EUAf-EUAs. (b) EUAf-DJIA. (c) EUAf-EURUSD. (d) EUAf-USDX. (e) EUAf-coal. (f) EUAf-oil. (g) EUAf-gas. (h) EUAf-electricity.

events sourced from the carbon market are essential risk factors that affect its homogeneous markets. The possible reason is that the price trend and market volatility of EUAf and EUAs are basically the same, the EUAf plays a guidance in the price discovery of the EUAs [8], and as a result, the trading risk and extreme risk of EUAf market can be easily transmitted to the EUAs market; this conclusion have proved in the study of Wang and Guo [1] that the European carbon future market implies more complex systemic risk than other carbon products markets. Additionally, the number of significance for the infected market of DJIA, oil, and electricity market are second only to significance number of EUAs. The number of significance for the EURUSD and coal is the smallest of all the market.

However, the number of significance for the infected market of capital market is generally equal to that of energy market. This shows that the asymmetric and extreme risk of carbon market can easily affect the capital market and energy market in rapid volatility trend although these two kinds of markets have different effects on dispersing carbon risk in the long run.

Further empirical results suggest that the contagion strength transmits from the EUAf to its infected market and is significantly different according to different volatility trend. Table 5 shows that the contagion of EUAf to EUAs in volatility rapid rise trend is higher for all the significant high-order moment channels than in volatility rapid decline trend. The possible reason is that the volatility rapid rise corresponds to more systematic risk (e.g., market asymmetric risk and extreme risk) which leads to a higher contagion coefficient than the volatility rapid decline trend. The conclusion is generally same with the research of risk contagion in the global stock market that was carried out by Forbes et al. and Renée et al. [3, 4]. The coskewness contagion of EUAf to DJIA in CS12 and CS21, EURUSD and electricity in CS21, and USDX in CS12 and CS21; contagion from EUAf to coal and gas in the channel of CS12; and contagion from EUAf to oil in CS12 and CS21 are higher in trend of volatility rapid rise than that of volatility rapid decline. The possible explanation is that the rapid rise of market volatility hides the possibility of continuous increasing of risk as risk averse agents prefer positive skewness to negative skewness [34]; investors will require more coskewness and expect to achieve more returns to offset the systemic risk. While the rapid decline of volatility means the declining of market risk, the lower market skewness can satisfy investors' expectations; therefore, the coskewness contagion coefficient in volatility rise rapidly trend is larger than that of decline trend.

Another evidence that convinces the stronger contagion in volatility rise rapidly trend concludes that the contagion coefficient from the sourced EUAf market to majority of its infected market in the channel of CK13 and CK31 are

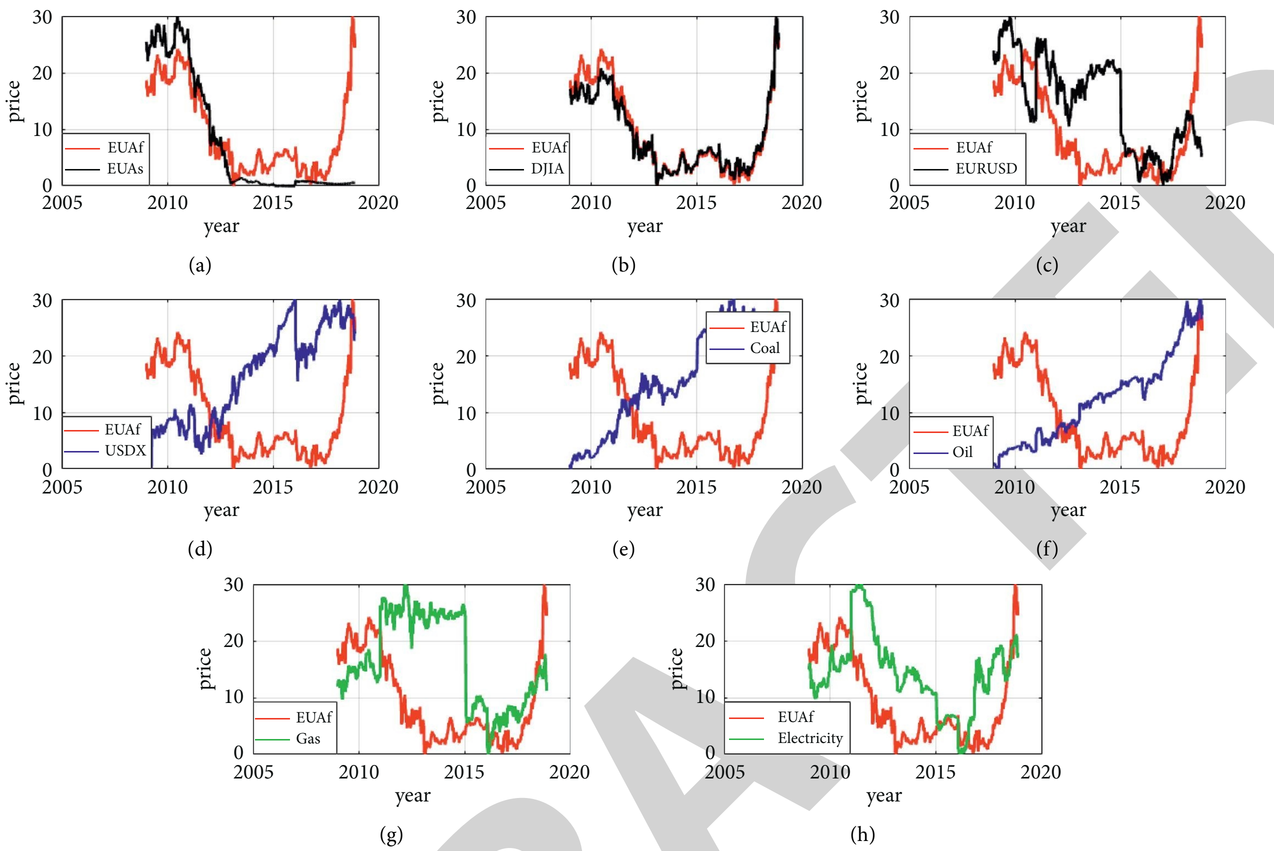


FIGURE 5: Standardized price change between carbon asset and its pricing factors in low volatility state (S3). (a) EUAf-EUAs. (b) EUAf-DJIA. (c) EUAf-EURUSD. (d) EUAf-USDX. (e) EUAf-coal. (f) EUAf-oil. (g) EUAf-gas. (h) EUAf-electricity.

TABLE 4: Descriptive statistics for the high-order moments coefficients of joint distribution between carbon market and its infected markets.

	State	EUAs	DJIX	EURUSD	USDX	Coal	Oil	Gas	Electricity
Coskewness12	S1	0.115	−0.011	−0.004	−0.049	0.124	−0.037	0.513	0.056
	S3	−0.141	0.063	−0.058	−0.061	−0.082	−0.078	−0.141	−0.306
	S2	−0.657	−0.805	−0.165	−0.342	−0.97	−0.439	−0.177	−0.367
Coskewness21	S1	0.111	0.085	−0.021	0.039	0.046	−0.045	0.008	−0.058
	S3	−0.084	−0.052	−0.042	0.047	−0.014	−0.041	−0.036	−0.057
	S2	−0.406	−0.026	−0.158	−0.103	0.014	−0.135	−0.032	−0.014
Cokurtosis13	S1	1.925	−0.236	0.675	−0.813	0.069	0.601	3.372	0.463
	S3	2.010	−1.028	0.298	−0.53	0.491	1.167	0.239	0.521
	S2	4.903	1.810	0.362	−0.011	0.917	3.392	1.276	2.594
Cokurtosis31	S1	1.604	0.009	0.483	−0.446	0.513	0.319	0.244	0.238
	S3	1.784	0.145	0.121	−0.211	0.163	0.285	0.373	0.352
	S2	3.421	0.451	1.188	1.101	0.113	0.578	0.591	0.437

generally higher in trend of volatility rise rapidly than in volatility decline rapidly. This can be explained by the higher systemic risk consistent with the volatility rise rapidly trend.

4.3.2. Analyzing the Risk Contagion in the Slow Volatility Trend. The empirical results showed in Table 6 that the number of significance in high-order moment contagion channels for the infected market of EUAs market is the highest among other infected markets when the market is in slow volatility trend. This conclusion is proved in above analysis. The number of significance for the infected market of DJIA,USDX, oil, gas, and electricity market are second only to significance number of EUAs. This indicates that asymmetric and extreme risks of the carbon market are more easily transmitted to these three markets rather than others during the slow volatility trend. When the carbon market is in slow volatility trend, there is less systemic risk than the rapid volatility trend; investors have enough time to develop portfolio strategies and increase

Table 5: Risk contagion in high-order moment channels from carbon market to its infected markets under the rapid volatility trend.

Carbon homogeneous market		Capital market			Energy market			
	EUAs	DJIA	EURUSD	USDX	Coal	Oil	Gas	Electricity
Panel A: rapid decline in volatility of EUAf (S2-S1)								
FR	2.85***	0.01	3.11	2.25	0.92	1.11	0.08**	0.08*
CS12	22.35***	19.88***	1.21	3.56***	6.1**	11.87***	24.78***	0.62
CS21	7.21***	0.03***	1.92***	2.08***	0.25	0.44***	0.01	0.47***
CK13	163.34***	10.95***	39.12	41.04	198.6***	289.9***	123.29	42.88***
CK31	87.59***	19.22***	21.81***	18.1***	36.9	85.18***	40.68***	44.45***
Panel B: rapid rise in volatility of EUAf (S1–S2)								
FR	15.5***	0.09	0.62	0.59	0.34	1.51	0.27**	0.16*
CS12	37.5***	21.87***	1.48	4.33***	7.05**	14.53***	30.23***	0.69
CS21	12.09***	0.04***	2.34***	2.53***	0.06	0.53***	0.01	0.53***
CK13	387.49***	11.67***	4.86	3.61	0.29***	293.5***	294.11	45.49***
CK31	222.49***	19.53***	26.24***	19.2***	0.44	114.3***	49.03***	46.31***

Note. S1, S2, and S3 denotes the stability, high volatility and low volatility state of the carbon market, respectively. The symbols *, **, and *** denote the statistical significance at the 10%, 5%, and 1%, respectively.

Table 6: Risk contagion in high-order moment channels from carbon market to its infected markets under the slow volatility trend.

Carbon homogeneous market		Capital market			Energy market			
	EUAs	DJIA	EURUSD	USDX	Coal	Oil	Gas	Electricity
Panel A: slowly decline in volatility of EUAf (S3–S1)								
FR	0.59***	0.02**	0.03**	0.41***	0.01***	0.02***	0.04***	0.09***
CS12	7.24***	32.29***	1.01	0.03**	11.13	0.48***	128.44	5.15**
CS21	3.71**	3.46	0.35	0.03	0.67	0.002	0.73	0.12
CK13	14.27***	3.45***	11.34***	11.9***	17.26***	70.1***	1.39***	0.02***
CK31	17.31**	9.07	11.62	14.54	2.54	15.17	8.39*	4.61
Panel B: slowly rise in volatility of EUAf (S1–S3)								
FR	10.72***	0.09**	0.61**	0.61***	0.33***	0.47***	0.27***	0.15***
CS12	10.25***	45.16***	1.12	0.03**	11.96	0.49***	135.63	5.52**
CS21	5.24**	4.26	0.39	0.03	0.73	0.004	0.77	2.11
CK13	30.88***	2.21***	11.21***	11.76***	43.59***	72.5***	1.64***	0.91***
CK31	36.36***	5.43	10.05	0.63	0.36	2.99	8.91*	2.001

Note. S1, S2, and S3 denotes the stability, high volatility, and low volatility state of the carbon market, respectively. The symbols *, **, and *** denote the statistical significance at the 10%, 5%, and 1%, respectively.

the possibility of rational investment behavior. As we know, the DJIA, which reflects the global macroeconomy, has become an important investment tool for carbon investors to diversify risks and obtain excess returns [6]. The global economy affects the trend of carbon price, the price volatility of oil market hides real arbitrage opportunities for the carbon market investors, and therefore, these markets are more likely to be affected by the risk contagion of high-order moment channels from the carbon market. Additionally, the number of significance for the infected market of EURUSD and coal market is the smallest of all the market.

Further empirical results suggest that the contagion strength is significantly different according to different volatility trend. Table 6 shows that the contagion strength in slow rise of market volatility is higher in majority of case than that of slow decline volatility for the high-order moment contagion channel of CS12, CS21, CK13, and CK31. The potential explanation is the slow rise of market volatility promotes the increasing of systemic risk, despite contrasted to the increase of risk in rapid volatility trend analyzed above, the slow rise of market volatility can still increase the portfolio risk impacted by extreme events and thus improves the risk contagion compared with the slow decline volatility. This conclusion is similar to the previous analysis that the contagion is stronger in volatility rise rapidly trend than decline rapidly trend.

4.4. Robustness Test of the Empirical Results. We now perform some simple variations on our basic analysis, with an eye toward checking robustness with respect to the empirical results above. We choose the subperiod of 2013.1.2–2018.11.14 as the sample span for conducting the robustness test following the same research steps and methods mentioned above. The reason is that since 2013, the European carbon emissions trading system (the world's largest carbon trading market) has entered the third development stage. Compared with the previous two stages, the scope of emission reduction entities has gradually

Table 7: Robustness test of risk contagion in high-order moment channels under the trend of volatility rise rapidly and decline rapidly, respectively.

Carbon homogeneous market		Capital market			Energy market			
EUAs	DJIA	STOXX	EURUSD	USDX	Coal	Oil	Gas	Electricity

	EUAs	DJIA	STOXX	EURUSD	USDX	Coal	Oil	Gas	Electricity
				Panel A: rapid decline in volatility of EUAf (S2-S1)					
FR	1.57***	0.05	0.05	1.92	0.23	0.016	0.12	0.22	0.03
CS12	12.66***	0.72	0.04	0.69	7.02	1.18***	0.72	19.17***	0.25
CS21	6.81***	0.98**	1.25***	0.001**	1.11***	0.37	0.13	0.08	0.51
CK13	254.41***	13.13*	0.02	16.71	8.49	62.2***	0.02	15.34	2.84
CK31	83.59***	16.32***	4.57***	5.47***	2.51***	0.62**	8.8***	2.81***	0.62***
				Panel B: rapid rise in volatility of EUAf (S1-S2)					
FR	5.65***	0.01	0.03	0.32	0.36	0.02	0.05	0.03	0.01
CS12	20.28***	0.75	0.04	0.79	7.96	1.19***	0.76	21.93***	0.26
CS21	10.89***	0.99**	1.25***	0.02**	1.25***	0.37	0.13	0.12	0.52
CK13	453.45***	4.84*	0.59	5.92	13.41	75.91**	6.96	74.13	0.06
CK31	143.43***	6.68***	8.96***	19.05***	26.2***	1.33***	13.2***	5.95***	1.34***

Note. S1, S2, and S3 denotes the stability, high volatility, and low volatility state of the carbon market, respectively. The symbols *, **, and *** denote the statistical significance at the 10%, 5%, and 1% respectively.

Table 8: Robustness test of risk contagion in high-order moment channels under the trend of volatility rise slowly and decline slowly, respectively.

	EUAs	DJIA	STOXX	EURUSD	USDX	Coal	Oil	Gas	Electricity
	Carbon homogeneous market		Capital market				Energy market		
				Panel A: slowly decline in volatility of EUAf (S3–S1)					
FR	0.03***	0.04	0.05	0.01	0.08	0.01	0.008	0.01	0.006
CS12	3.59***	1.66	0.09	3.55**	7.56	5.26***	19.3***	79.35	2.17***
CS21	0.52***	0.26*	0.56	0.86*	0.57	0.39	0.46	0.64	0.35
CK13	230.6***	3.53*	1.86***	0.11***	1.93*	24.91***	8.13***	650.72	168.43*
CK31	12.9***	5.11*	0.67*	9.26	6.52	0.21**	2.64***	0.53**	0.87
				Panel B: slowly rise in volatility of EUAf (S1–S3)					
FR	4.61***	0.04	0.006	0.31	0.35	0.01	0.01	0.03	0.01
CS12	5.08***	1.66	0.09	3.55**	7.69	5.45***	19.4***	79.73	2.23***
CS21	0.73***	0.25*	1.87	0.85*	0.59	0.41	0.39	0.67	0.34
CK13	335.4***	4.82***	3.95***	4.99***	16.59*	27.97***	12.8***	725.11	182.19*
CK31	14.7***	7.58*	0.72*	0.41	0.004	0.23***	2.69***	0.36**	2.06

Note: S1, S2, and S3 denotes the stability, high volatility, and low volatility state of the carbon market, respectively. The symbols *, **, and *** denote the statistical significance at the 10%, 5%, and 1%, respectively.

expanded, the trading products have gradually enriched, and especially the proportion of auctions in the process of quota allocation has gradually increased, while the allocation of free quotas is decreasing.

Tables 7 and 8 illustrate the robustness test of risk contagion from the sourced carbon market to its infected markets in high-order moment channels. The main findings are as follows.

Firstly, there is a significant risk contagion effect from the carbon market to its infected markets in majority of the high-order moment channels rather than the FR contagion in the low-order perspective both in rapid and slow volatility trend. Secondly, the number of significance in the high-order moment channels is the largest compared with other infected markets as for the EUAs both in rapid and slow volatility trend. Thirdly, the contagion strength of carbon market to its infected market in majority of high-order moment channel is higher in volatility rapid and slow rise trend than that of decline trend. Generally, the robustness result of risk contagion is essentially consistent with the conclusion. This demonstrates that the conclusions of this article based on the proposed model are reliable.

5. Conclusions

Research on the risk contagion between carbon market and its infected markets can not only reveal the risk contagion direction and strength but also provide reference for investors. However, the existing studies mainly focus on exploring information linkage and volatility spillover from the perspective of low-order moment attributes of returns. For remedying the defects of the existing research, the contribution of this article is to reveal the significant high-order moment contagion channel and strength of carbon market to its infected market under different market volatility trends and explore the contagious difference caused by volatility trends. Based on the empirical results, some main conclusions are obtained as follows.

Firstly, the contagion of the carbon market to its infected markets happens mostly in the high-order moment channel, rather than the view of the low-order moment channel in previous studies. Secondly, the market of EUAs, DJIA, coal, and oil are more likely to trigger risk contagion in the carbon market.

Thirdly, the contagion strength of EUAf to its infected market is higher in the majority of high-order moment channels than in the slow volatility trend. The possible explanation is the rapid volatility of market trends may hide more systemic risks and uncertainties; the contagion power is higher than the slowly volatility trend.

The methodology measured the risk contagion from the view of high-order moment has potential advantage than other models. Moreover, the methodology makes a significant judgment on the existence of risk contagion before determining the risk contagion coefficient; this measure is consistent with the connotation of risk contagion proposed by King and Wadhwani [35]. However, frankly, this methodology is designed to study contagion relationships only between two assets or two markets; it's unsuitable to study contagion relationships among the capital market, the energy market, and the carbon market at the same time. Therefore, the methodology and the risk contagion model can be improved in the future study to be more suitable for the high-order moment channel risk contagion among the multimarkets case.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work has been fully supported by Key Projects of Humanities and Social Sciences in Anhui Province of China (SK2021A1105); the Domestic Visit and Training Program for Outstanding Young Teachers in Anhui Province of China (gxgnfx2021098); the Youth Fund Project for Humanities and Social Sciences of the Ministry of Education of China (21YJC790152); and the research project of Humanities and Social Sciences of Anhui Vocational College of Finance and Trade (tzpyxj123).

References

[1] Y. Wang and Z. Guo, "The dynamic spillover between carbon and energy markets: new evidence," *Energy*, vol. 149, pp. 24–33, 2018.

[2] K. G Andrew, "Does international financial contagion really exist?" *International Finance*, vol. 6, no. 2, pp. 179–199, 2003.

[3] K. J. Forbes and R. Rigobon, "No contagion, only interdependence: measuring stock market comovements," *The Journal of Finance*, vol. 57, no. 5, pp. 2223–2261, 2002.

[4] F. Renée, V. L. Martin, and C. Tang, "A new class of tests of contagion with applications," *Journal of Business & Economic Statistics*, vol. 28, no. 3, pp. 423–437, 2010.

[5] J. Li, C. Liang, X. Zhu, X. Sun, and D. Wu, "Risk contagion in Chinese banking industry: a transfer entropy-based analysis," *Entropy*, vol. 15, no. 12, pp. 5549–5564, 2013.

[6] Y.-J. Zhang and Y.-M. Wei, "An overview of current research on EU ETS: evidence from its operating mechanism and economic effect," *Applied Energy*, vol. 87, no. 6, pp. 1804–1814, 2010.

[7] B. Y. Chang, P. Christoffersen, and K. Jacobs, "Market skewness risk and the cross section of stock returns," *Journal of Financial Economics*, vol. 107, no. 1, pp. 46–68, 2013.

[8] J. Chevallier, "Volatility forecasting of carbon prices using factor models," *Economics Bulletin*, vol. 30, no. 2, pp. 1642–1660, 2010.

[9] M. Mansanet-Bataller, J. Chevallier, M. Hervé-Mignucci, and E. Alberola, "EUA and sCER phase II price drivers: u," *Energy Policy*, vol. 39, no. 3, pp. 1056–1069, 2011.

[10] J. Chevallier, "Global imbalances, cross-market linkages, and the financial crisis: a multivariate Markov-switching analysis," *Economic Modelling*, vol. 29, no. 3, pp. 943–973, 2012.

[11] M. Balcılar, R. Demirer, and S. Hammoudeh, "Risk spillovers across the energy and carbon markets and hedging strategies for carbon risk," *Energy Economics*, vol. 54, pp. 159–172, 2016.

[12] B. J. Lutz, U. Pigorsch, and W. Rotfuß, "Nonlinearity in cap-and-trade systems: the EUA price and its fundamentals," *Energy Economics*, vol. 40, pp. 222–232, 2013.

[13] V. Medina and A. Pardo, "Is the EUA a new asset class?" *Quantitative Finance*, vol. 13, no. 4, pp. 637–653, 2013.

[14] S. Hammoudeh, A. Lahiani, D. K. Nguyen, and R. M. Sousa, "An empirical analysis of energy cost pass-through to CO 2 emission prices," *Energy Economics*, vol. 49, pp. 149–156, 2015.

[15] J. C. Reboredo, "Volatility spillovers between the oil market and the European Union carbon emission market," *Economic Modelling*, vol. 36, pp. 229–234, 2014.

[16] C. Zhang, Y. Yang, and T. Zhang, "The integrated measurement about carbon finance market risk of commercial banks based on copula model," *Chinese Journal of Management Science*, vol. 23, no. 4, pp. 61–69, 2015.

[17] N. Yuan and L. Yang, "Asymmetric risk spillover between financial market uncertainty and the carbon market: a GAS-DCS-copula approach," *Journal of Cleaner Production*, vol. 259, Article ID 120750, 2020.

[18] X. Sun, W. Fang, X. Gao, H. An, S. Liu, and T. Wu, "Complex causalities between the carbon market and the stock markets for energy intensive industries in China," *International Review of Economics & Finance*, vol. 78, pp. 404–417, 2022.

[19] X. Tan, K. Sirichand, A. Vivian, and X. Wang, "How connected is the carbon market to energy and financial markets? A systematic analysis of spillovers and dynamics," *Energy Economics*, vol. 90, Article ID 104870, 2020.

[20] Y. Hao, C. Tian, and C. Wu, "Modelling of carbon price in two real carbon trading markets," *Journal of Cleaner Production*, vol. 244, Article ID 118556, 2020.

[21] F. Nazifi and G. Milunovich, "Measuring the impact of carbon allowance trading on energy prices," *Energy & Environment*, vol. 21, no. 5, pp. 367–383, 2010.

[22] P. Aatola, M. Ollikainen, and A. Toppinen, "Price determination in the EU ETS market: theory and econometric

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9805150, 1 page
https://doi.org/10.1155/2022/9805150

Retraction

Retracted: Risk Assessment of Agricultural Economic Management Based on the Multivariate Statistical Computing Method

Security and Communication Networks

Received 12 November 2022; Accepted 12 November 2022; Published 23 November 2022

Security and Communication Networks has retracted the article titled "Risk Assessment of Agricultural Economic Management Based on the Multivariate Statistical Computing Method" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The author does not agree to the retraction.

References

[1] Y. Zhou, "Risk Assessment of Agricultural Economic Management Based on the Multivariate Statistical Computing Method," *Security and Communication Networks*, vol. 2022, Article ID 8547306, 11 pages, 2022.
[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8547306, 11 pages
https://doi.org/10.1155/2022/8547306

Research Article

Risk Assessment of Agricultural Economic Management Based on the Multivariate Statistical Computing Method

Yan Zhou

College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, Guangdong Province, China

Correspondence should be addressed to Yan Zhou; zhouyan1980@scau.edu.cn

Received 13 January 2022; Revised 28 February 2022; Accepted 11 March 2022; Published 14 April 2022

Academic Editor: Chin-Ling Chen

The modernization process of Chinese agriculture has posed new challenges to agriculture economic management. However, existing studies focus on financial and ecological and environmental risks of agriculture economic management while lacking the necessary attention to other types of agricultural economic management. Therefore, we first propose that the risk of agricultural economic management is of five types—economic, social, political, cultural, and ecological and environmental risks—and further clarify the interactions among the five risk types. Given that the five types of risks are nested with each other, we adopted a multivariate statistical algorithm based on complex network theory to scientifically evaluate the risk management of agriculture economy. The results show the applicability of the algorithm to risk clustering analysis and risk coefficient estimation. The article concludes with the corresponding theoretical and practical implications.

1. Introduction

The high attention of the Chinese government has greatly promoted the modernization and transformation of Chinese agriculture, presenting a sea of changes with respect to increasing the agricultural output value, improving agricultural production conditions, improving the new agricultural business system, and forming new dynamics of agricultural development, which have made China leap to the forefront of the world in agricultural opening achievements and made rural residents' life move toward an overall well-off position [1]. One of the important reasons for the success of agricultural modernization is the market-oriented transformation of agricultural economic management. Although the transformation is a proactive choice to adapt to the new situation of international comprehensive national power competition in the post-financial crisis era [2], the market-oriented transformation of agriculture also brings huge new risks and challenges. However, the academic community focuses on the financial risks and ecological and environmental risks of agriculture while paying less attention to other risks of agricultural economic management [3, 4], and it also lacks a systematic classification of the types of risks of agricultural economic management [5], which leads to the insufficient explanatory power of theories on the anti-risk capacity of China's agricultural economy; in addition, it is necessary to adopt a scientific approach to the types and structures of agricultural economic management risks and their related relationships. Systematic analysis is necessary to provide reference for the risk response strategy of China's agricultural economic management.

In fact, after China's accession to the World Trade Organization, the pace of agricultural opening to the outside world has accelerated significantly, and it has become a major country in the world in terms of agricultural opening to the outside world. In the more open global market and agricultural competition, competitiveness has become a hot spot in China's agricultural economic management [6]. In an open market environment, the ability of agricultural production operators to produce agricultural products that outperform similar or alternative products in a given market in terms of buyer value, thus battling to win and maintain market share in competition and winning profits for agricultural production operators, is agricultural market competitiveness [7, 8]. The international competitiveness of agricultural products consists of three dimensions, namely, price competitiveness, quality competitiveness, and reputation competitiveness, which are organically constituted [9]. From the perspective of the international market, the all-round competition based on quality and technological content replaces the pure price competition [10]. To improve the competitiveness of China's agricultural products, the focus is on improving the quality of agricultural products and fully transforming the comparative advantages of agricultural products into competitive advantages [11]. Among them, industrialization is an important way to improve the competitiveness of agriculture. How to determine the focus of support in the negotiation process of China's accession to the WTO and how to coordinate the role of financial expenditures and other resources, including positions and allies in international multilateral negotiations, to improve the efficiency of government support and the comprehensive competitiveness of agricultural products are of great significance [12, 13]. On the other hand, in the context of world agricultural and rural development dynamics, agriculture faces the pressure and challenges of mitigating and adapting to climate change and feeding a large number of people, and the promotion of sustainable agriculture is an important means of adapting to climate change, reducing greenhouse gas emissions from agriculture and mitigating deforestation, among other issues. Climate [14], climate change [15], agricultural adaptation [16], and crop modeling [17] have become hot topics in international agricultural economics research. The research frontier of agricultural economic management in China is also in line with the international agricultural economics research lineage and evolutionary trends [18]. Climate change, mainly characterized by rising temperatures, is a serious challenge common to all countries in the world today [19], and most farmers are able to recognize the phenomenon of climate change and the impact of climate change on agricultural production [20]. Among them, farmers' age [21], education level [22], income level [23], and their level of awareness of climate change [24] have a significant impact on whether they adopt adaptive behaviors. In addition, log production function models have been widely used to analyze the impact of climate change on the yield of major food crops [25], emphasizing the introduction of more active and effective climate policies to effectively mitigate the adverse effects of climate change on food production in China [26]. Obviously, the current research focuses on the financial and environmental risks of agricultural economic management, and the relevant research system is fragmented and lacks a comprehensive assessment of the risks of agriculture economic management. More importantly, risk evaluation still adopts traditional hierarchical analysis or factor analysis while ignoring the complex structure of agricultural economic management risks, which urgently needs to be deconstructed by applying the theory of complex networks. Therefore, this article first analyzes the types of risks in agriculture economic management and the logical relationship between each type in depth, and it adopts the factor analysis method based on the complex network analysis theory to make a comprehensive assessment of the risks in agriculture economic management in order to provide theoretical references for risk management countermeasures in agricultural economic management.

This article is mainly composed of five sections. The first section introduces the importance of scientifically evaluating the risks of agricultural economic management. The second section identifies and introduces five risk factors in agricultural economic management, including economic, social, political, cultural, and ecological and environmental risks. The third section explains a multivariate statistical algorithm based on a complex network. The fourth section provides the data analysis results. The fifth section discusses the theoretical and practical contribution of this article.

2. Identification of Risk Factors in Agricultural Economic Management

In contrast to previous studies that focused solely on the economic risks and the environmental risks of agriculture, we argue that agriculture economic management has various types of risks, which can be roughly divided into five types according to different causes: economic risks, social risks, political risks, cultural risks, and ecological and environmental risks.

Economic risks. The main tasks of agriculture economic management are to reduce production capacity in agriculture through structural adjustment, to reduce the cost of corn planting in the "sickle curve" area, and to make up for the shortcomings. According to the Ministry of Agriculture's guidelines on "de-capacity," optimizing the structure and appropriately reducing corn production capacity are the top priority of the current reforms on agriculture. The cost of removing production capacity requires subsidies for fallowing, subsidies for credit guarantee costs, and subsidies for science and technology innovation. The paths to reduce agricultural production costs include moderate scale operation, development of agricultural science and technology, and improvement of agricultural infrastructure; the costs required to make up for the shortcomings of agricultural development include those of endogenous factors. Therefore, from the perspective of the three tasks of "removing production capacity," reducing costs, and making up for shortcomings, agriculture economic management requires a considerable amount of cost expenditure [27].

Against the background of a further slowdown in international economic growth and a domestic economic situation that is "stabilizing and improving" but with difficulties and challenges, the government, as the main bearer of the costs of agriculture management, will incur huge pressure on its fiscal balance to eliminate the huge reform costs. At the same time, increasing the share of fiscal expenditure on agriculture economic management will inevitably reduce fiscal investment expenditure in other areas, which will lead to a certain degree of shrinkage of production capacity in other areas, thus reducing the total output level of the society as a whole. In the case that farmers' production skills have not yet been transformed and the scope of production and operation has not been completely adjusted, crop rotation and fallow required by the agriculture economic management and the short-term break in industrial transformation caused by "grain to feed" may also affect the development of rural economy [28].

Social risks. Agriculture economic management also involves certain social risks. First, the adjustment of agricultural production and management structure and regional production structure is also the process of interest structure adjustment, and the change in the interest pattern will inevitably lead to social contradiction and even conflicts. Second, agriculture economic management requires large-scale agricultural operation, which is premised on the transfer of agricultural land management rights, and the transfer of agricultural land management rights also entails a series of risks, such as land loss, unemployment, loss of livelihood security, division of rural areas between two classes, and damage to farmers' rights and interests. In addition, the cost-sharing mechanism of agricultural economic management has not yet been formed, the responsibility and rights of sharing subjects are not divided, and the costs are unevenly shared among classes, regions, and urban–rural areas, which may lead to increased conflicts among classes, regions, and urban–rural areas, thus causing social risks [29].

Political risk. Agriculture economic management involves international and domestic political risks. The international political risk of agricultural economic management refers to the manipulation of international food market prices by some big countries, which makes China's food production and import and export subject to the control of others, such as the entry of a large amount of foreign investment capital into the food market to buy and sell short and hoard, thus causing the risk of instability of the national regime. First, agriculture economic management requires large-scale transfer of agricultural land to accommodate large-scale operation, but "if important areas such as rural contracted land are controlled by private capital, the economic foundation of socialism in China will no longer exist," and the nature of the socialist state will be challenged. Second, if agriculture economic management leads to a slowdown in rural economic development, an increase in unemployment, and social class confrontation, the rural society will become unstable. Furthermore, as agriculture economic management progresses, farmers' property income, compensatory income, and transfer income will increase, but this will also provide opportunities for corruption among village cadres, which will lead to confrontation between villagers and village cadres, strain relations between the cadres and the villagers, and affect the party's ruling base in rural areas. Finally, there is a risk that the transfer of agricultural land management rights will weaken the functioning of villagers' self-governing organizations and weaken the party's ruling base in rural areas. The transfer of agricultural land intensifies the tendency of individual villagers to decentralize, and the growing power of local clans and families will affect the authority of grass-roots organizations, which will reduce the villagers' self-governing organizations to mere institutions responsible for handling daily village affairs, thus weakening the ruling party's foundation in rural areas [30].

Cultural risks. Any nation contains both urban and rural cultures, and the different cultures of urban and rural areas are both complementary and conflicting. In the pre-modern agricultural society, the Chinese countryside had strong stability and was in an almost balanced and stable state, which was formed thanks to the long-standing rural culture. However, during the agriculture economic management process, along with the gradual penetration of urban industrial and commercial capital and urban culture into the countryside, the traditional local, human, and acquaintance society in the Chinese countryside has been reduced due to living customs and cultural background. The traditional local, human, and acquaintance society in rural China will be challenged by the differences in living customs and cultural backgrounds, and rural production lifestyles, value pursuits, family values, and ethics will be changed accordingly. In this context, individualism and money worship, lack of social integrity and faith, and moral crisis may occur in the rural society, and the farming culture will gradually disappear [31].

Ecological and environmental risks. Developing ecological agriculture and protecting the ecological environment are two of the main objectives of China's agricultural economic management. As an agricultural production method that follows the laws of ecological economy and is closely related to the reality of Chinese agriculture, ecological agriculture will become an effective way to achieve reforms in Chinese agriculture. However, agriculture economic management poses ecological and environmental risks. First, the nonagricultural use of agricultural land will threaten the rural ecological environment. In the process of land transfer, a considerable amount of public land is transformed into nonagricultural construction land, and nonagricultural use of agricultural land will inevitably bring a series of environmental pollution and ecological damage, such as air pollution, water pollution, and noise pollution. Second, once the agricultural land is not properly organized and technically managed, the stability of the agricultural ecosystem will be damaged, and the productivity and soil properties of the agricultural land will be affected, because of which land degradation will be inevitable. Finally, large-scale agricultural management will bring great threats to the ecological environment. The scale operation of agriculture destroys biodiversity, and the excessive reliance on and use of synthetic chemicals have laid hidden dangers on human food and water safety; the scale operation weakens or even breaks the material–energy cycle between agriculture and nature, completely rejecting the natural succession of biological communities and self-regulation within the tolerance limit and breaking the local microcirculation of agriculture [32].

Although agricultural economic management is a change in the agricultural field, due to its comprehensiveness and complexity and the interconnectedness of social systems, the five types of risks are not isolated from each other in a specific environment, but affect each other, are contagious, reinforce each other, and overlap with each other. When impacted by external or internal contingent events, each risk point will resonate and link up under the influence of the domino effect and eventually evolve into systemic risk (see Figure 1), and its impact will go far beyond the agricultural sector itself. For example, in order to optimize the

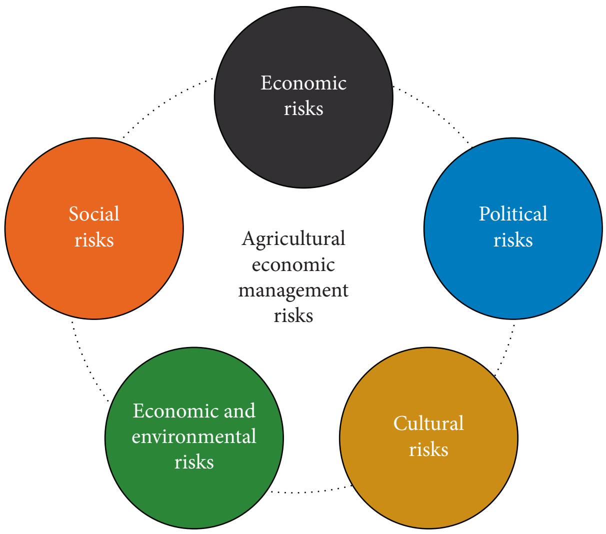


Figure 1: Agricultural economic management risks.

regional industrial structure and protect the ecological environment, the government must increase the subsidies for farmers in ecological functional areas when designating ecological functional areas and implementing the fallow crop rotation system, while under the hard constraint of government fiscal expenditure, the increase in agricultural expenditure will inevitably reduce fiscal expenditure in other areas, and the reduction in fiscal expenditure in other areas will affect their employment absorption capacity, resulting in an increase in social unemployment and instability. In addition, if the ecological function zones are designated to strictly prohibit certain types of agricultural production and operation activities and the fallow rotation system is implemented, the number of structurally unemployed farmers will increase, and the income of rural residents will decrease; in addition, if the fallow subsidies are not fully provided or misappropriated, farmers' dissatisfaction will accumulate, and grievances against the government will be formed. In this case, the increase in unemployment not only directly affects the speed of economic development but also affects the stability of rural areas and the society as a whole. In addition, if the scale of agricultural land transfer is too large and too fast, the large-scale operation will replace the small farmers' economy too quickly; especially, if most of the rural land is controlled by private capital, the villagers' autonomy will be reduced, and the party's ruling foundation in rural areas will be shaken, thus creating political risks. In addition, if the government does not fully deal with the employment issue of farmland transfer, the large-scale transfer of farmland by the government will inevitably lead to the increase in rural unemployment and the loss of farmers' interests.

3. Methodology

Traditional risk assessment methods generally construct the evaluation index system first and then model it using hierarchical analysis and other methods, but given that the risks of agricultural economic management in fact constitute a complex network structure, the individual risk evaluation factors interact with each other and are difficult to be stripped away. Therefore, based on complex network theory, we adopted the factor analysis algorithm for risk modeling of agricultural economic management. Importantly, the association structure in a complex network is a collection of several network nodes, and the edges between the nodes within the collection are dense while the edges between the collections are relatively sparse. Complex network association structure mining can clearly and accurately characterize the topology of network structure, help reveal the functional characteristics of each dimension of complex systems, understand the group characteristics of complex networks, and scientifically evaluate abstract models [33, 34]. Specifically, the method starts from the correlation matrix of many observed variables and groups the observed variables according to the magnitude of the correlation so that the correlation between observed variables within the same group is high and the correlation between variables in different groups is low (Figure 2). Each group of variables can be represented by an unobservable implicit variable, called the common factor, which acts on all variables. On this basis, the original variables are decomposed into a sum of two parts, one representing a linear combination of a few unobservable implied variables and the other a special factor, which is unrelated to the common factor and only correlates with the original variables themselves.

There are n samples, each with p observations. These p observations can be expressed as p components of a random vector $X = (X_1, X_2, \ldots, X_p)^T$ after normalization. Let the mean vector $E(X) = 0$ of this random vector X and the covariance matrix $\operatorname{cov}(X) = \Sigma$ be equal to the correlation array of X. $F = (F_1, F_2, \ldots, F_m)^T$ $(m < p)$, which denotes the m common factors with mean vector $E(F) = 0$. The covariance matrix $\operatorname{cov}(F) = \mathrm{In}$, where In denotes the unit diagonal array. $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T$ is called the special factor, which is only related to the components X_i $(i = 1, 2, \ldots, p)$ of the random vector X, is independent of each other and F, and has $E(\varepsilon) = 0$; let the components of ε be independent of each other; the covariance matrix $\Sigma\varepsilon$ is given as follows:

$$\operatorname{cov}(\varepsilon) = \sum_{\varepsilon} = \begin{bmatrix} \sigma_{11}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{pp}^2 \end{bmatrix}. \tag{1}$$

Based on the above-mentioned description, the following model is defined as a factor model:

$$\begin{cases} X_1 = k_{11}F_1 + k_{12}F_1 + \cdots + k_{1m}F_m + \varepsilon_1, \\ X_1 = k_{21}F_1 + k_{22}F_1 + \cdots + k_{2m}F_m + \varepsilon_2, \\ \cdots\cdots \\ X_p = k_{p1}F_1 + k_{p2}F_1 + \cdots + k_{pm}F_m + \varepsilon_p, \end{cases} \tag{2}$$

where $k = (k_{ij})_{pxm}$ is the coefficient matrix, usually called the factor loading matrix. The larger the absolute value of each factor loading k_{ij}, the greater the correlation between X_i and F_j. Usually, there is $|k_{ij}| \le 1$. According to equation (2), the covariance between the variables X_i and F_j is as follows:

$$\operatorname{cov}\left(X_i, F_j\right) = \operatorname{cov}\left(\sum_{j=1}^{m} k_{ij} F_j + \varepsilon_i, F_j\right). \tag{3}$$

Since $F_1, F_2, \ldots, F_m$ are independent of each other and $F_1, F_2, \ldots, F_m$ and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ are also independent of each other, it follows that

$$\operatorname{cov}\left(X_i, F_j\right) = k_{ij}. \tag{4}$$

That is to say, the factor loading k_{ij} indicates the degree of correlation between X_i and F_j. According to equation (2), the correlation coefficient between the variables X_i and X_j is as follows:

$$r_{ij} = k_{i1}k_{j1} + k_{i2}k_{j2} + \cdots + k_{im}k_{jm}. \tag{5}$$

Equation (5) shows that the correlation coefficient between X_i and X_j is also larger when both variables X_i and X_j have larger loadings on a common factor.

For a complex network graph G with n nodes, let its adjacency matrix be A. Let $X = (X_1, X_2, \ldots, X_n)$ be a random vector, where each component $X_i (i = 1, 2, \ldots, n)$ represents the weights of the edges between node i and other nodes in the network, and n denotes the weights of the edges of node i with other nodes in the network. Based on this definition, the n components of the i-th row vector a_i of the adjacency matrix A can be regarded as the n sampled values of the random variable X_i corresponding to node i. In a complex network, the association structure is a division of the set of nodes in the network, and each subset of nodes is called an association. The edges between nodes belonging to the same association are tightly connected, while the edges between nodes belonging to different associations are relatively sparse. This definition implies the fact that if node i and node j belong to the same association in a complex network G, the components of the corresponding row vectors a_i and a_j of nodes i and j in A are relatively similar. Since the row vectors a_i and a_j can be regarded as vectors of n sampled values of random variables X_i and X_j, respectively, the distributions of random variables X_i and X_j are similar, so it is known that there is a large correlation between X_i and X_j. Based on the above-mentioned analysis, if the nodes i and j belong to the same association, the correlation between the random variables X_i and X_j in the complex network G is larger.

According to the theory of factor analysis, n random variables $X_i (i = 1, 2, \ldots, n)$ with large correlations with each other can be linearly represented by m common factors F_i $(i = 1, 2, \ldots, m)$. Each common factor reflects a set of random variables with large correlation. The specific formula is shown as follows:

$$\begin{cases} X_1 = k_{11}F_1 + k_{12}F_1 + \cdots + k_{1m}F_m + \varepsilon_1, \\ X_1 = k_{21}F_1 + k_{22}F_1 + \cdots + k_{2m}F_m + \varepsilon_2, \\ \qquad \cdots\cdots \\ X_n = k_{n1}F_1 + k_{n2}F_1 + \cdots + k_{nm}F_m + \varepsilon_n, \end{cases} \tag{6}$$

where a_{ti} denotes the components of the adjacency matrix A of the network graph G. $\overline{a_i} = \sum_{t=1}^{n} a_{\mathrm{it}}/n$ denotes the mean of the elements in the i-th row vector of the matrix A.

$$r_{ij} = \frac{n-1}{n} \frac{\sum_{i=1}^{n} \left(a_{ti} - \overline{a_i}\right)\left(a_{tj} - \overline{a_j}\right)}{\sqrt{\sum_{i=1}^{n} \left(a_{ti} - \overline{a_i}\right)^2} \sqrt{\sum_{t=1}^{n} \left(a_{tj} - \overline{a_j}\right)^2}.} \tag{7}$$

Given that matrix R is most similar to matrix $R = (r_{ij})_{n\times n}$, the elements of the two matrices R and R' are obtained from equations (7) and (8), respectively.

$$R'' = R + 1 = \begin{bmatrix} r_{11} + 1 & \ldots & r_{1n} + 1 \\ r_{21} + 1 & \ldots & r_{2n} + 1 \\ & \ldots & \\ r_{n1} + 1 & \ldots & r_{nn} + 1 \end{bmatrix}. \tag{8}$$

Since equation (8) is the Pearson correlation coefficient formula and each variable on the right-hand side of the equation is greater than zero, we have $|r_{ij}| \leq 1$ $(i, j \in [1, n])$, from which we know that element $r_{ij}{}'' \in [0, 2]$ in R''. By normalizing each row of R'', we get the following:

$$R''' = \frac{R''}{2}. \tag{9}$$

The transformation from matrix R' to R'' is linear and does not change the relative magnitudes of the components. The optimization problem (7) is approximated by the above-mentioned mathematical transformations into an optimization problem with nonnegative matrix decomposition, i.e., matrix R' is decomposed into the product of matrices K and K^T such that $||R''' - R'||_2$ is minimized and satisfies $(k_i, k_j) = 1$, $i, j \in [1, n]$, where k_i and k_j denote the row vectors of the i-th and j-th rows of matrix K, respectively. Currently, there are many techniques to decompose the above-mentioned matrices to obtain the affiliation matrices of the network nodes to each association structure and then identify the association structures existing in the network. Although the above-mentioned algorithms can mine the association structure, the characteristics of non-negative matrix decomposition algorithms determine that these algorithms can only obtain the association affiliation information of each network node and cannot further sense the hierarchical association structure in the network. Therefore, it cannot reveal the hierarchical organization among nodes in many real complex networks. In the subsequent subsections of this chapter, we improve the existing association splitting and clustering algorithms based on the factor analysis modeling of network association structure to realize the mining of hierarchical association structure in complex networks.

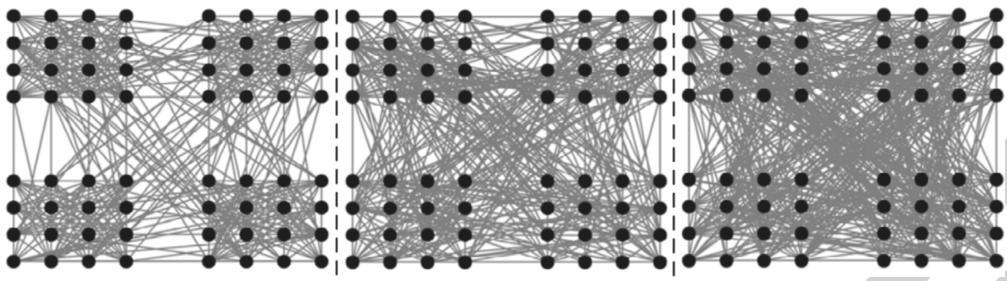


Figure 2: Schematic diagram of complex network.

A local minimum edge is defined as the set of edges in a complex network graph G that satisfy the following conditions:

$$r_{ij} = \min\left\{r_{xy} | (x = i, y\varepsilon\tau(i) - \{i\}) \cap (x = j, y\varepsilon\tau(j) - \{j\}\right). \tag{10}$$

$\tau(i)$ denotes the set of all neighboring nodes of node i. According to the definition of equation (10), the edges in this set are usually not directly adjacent to each other, and their weights are the minimum of the set of locally contiguous edges. A small network with several nodes is illustrated in Figure 3. The network has a typical two-association structure, with nodes of different colors. The weights of the connected edges between the nodes are similarities calculated by equation (8). According to equation (10), the set of local minimum edges in the current network topology can be found, and each local minimum edge is represented by a thick black line. It can be seen that two local minimum edges are not adjacent to each other and the influence of removing two edges at the same time is small.

In a complex network with a typical association structure, if removing an edge in the set of local minimal edges can increase the similarity of the edges between the nodes within the association connected to that edge, then selecting that edge as the edge to be removed can guarantee the association structure in the network with a higher probability of detection by the association splitting algorithm. In order to explore the mathematical characteristics of such edges, Figure 4 shows a part of the network structure graph truncated from the general complex network. This part of the network graph has a typical association structure, where the solid lines show the actual edges between the nodes of the network and the dashed lines show the edges between the nodes and the unintercepted part of the network and ignore the irrelevant details.

For a weighted network graph G, its adjacency matrix is A. Let $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ denote the row vectors of nodes X and Y in the adjacency matrix A of the network. $\mathrm{Var}(X)$ denotes the variance of the random vector corresponding to node X, and $\mathrm{cov}(X)$ denotes its covariance coefficient. Using the above-mentioned symbolic definitions, the covariance coefficients between nodes X and Y change after removing the continuous edge between X and K as follows:

$$\begin{aligned}
&\mathrm{cov}(X,Y) - \mathrm{cov}(X',Y) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) - \frac{1}{n}\sum_{i=1}^{n}(x_i' - \overline{x})(y_i - \overline{y}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i - x_i' - \overline{x} + \overline{x'}\right)(y_i - \overline{y}) \\
&= \frac{1}{n}(y_k - \overline{y})x_k.
\end{aligned} \tag{11}$$

Removing the edges between X and K increases the value of $\mathrm{cov}(X, Y)$ when $y_k = 0$. The $y_k = 0$ in Figure 5 indicates that there is no network edge between nodes Y and K. The variance between nodes X and Y changes after removing the edges between X and K as follows:

$$R_{XY} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}. \tag{12}$$

Overall, the specific workflow for the application of our multivariate statistical algorithm in risk assessment of agriculture economic management is shown in Figure 6. The algorithm models the relationship between network nodes and associations by factor analysis, and based on this, the similarity between nodes is calculated as the weight of the connected edges of the network with the help of the formula. The algorithm finds all the local minimum edges that satisfy the conditions in the current network topology and removes them. If all the local minimum edges in the current network do not satisfy the condition, all the found edges are deleted at once. The above-mentioned process is repeated until the algorithm finds the optimal association structure. In order to evaluate the merit of the current association structure in the network and stop the iterative process of the algorithm, the proposed algorithm in this section uses a similarity-based modularity formula. This formula compensates for the shortcomings of the classical modularity definition and enables a more accurate evaluation of the association classification results.

4. Results and Discussion

Figure 7 shows the experimental results of the factor analysis–based association splitting algorithm proposed in this section for this network graph. From the information labeled

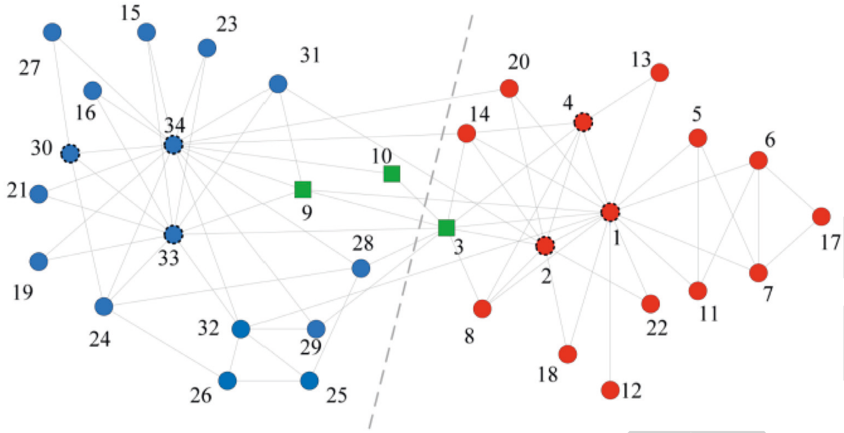
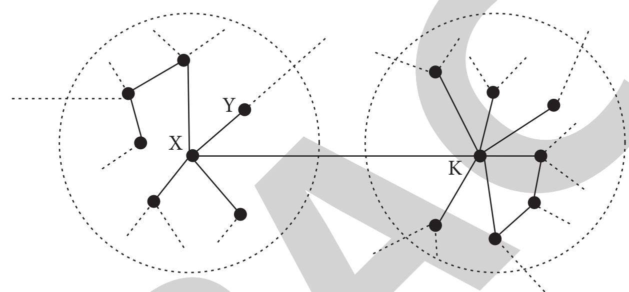
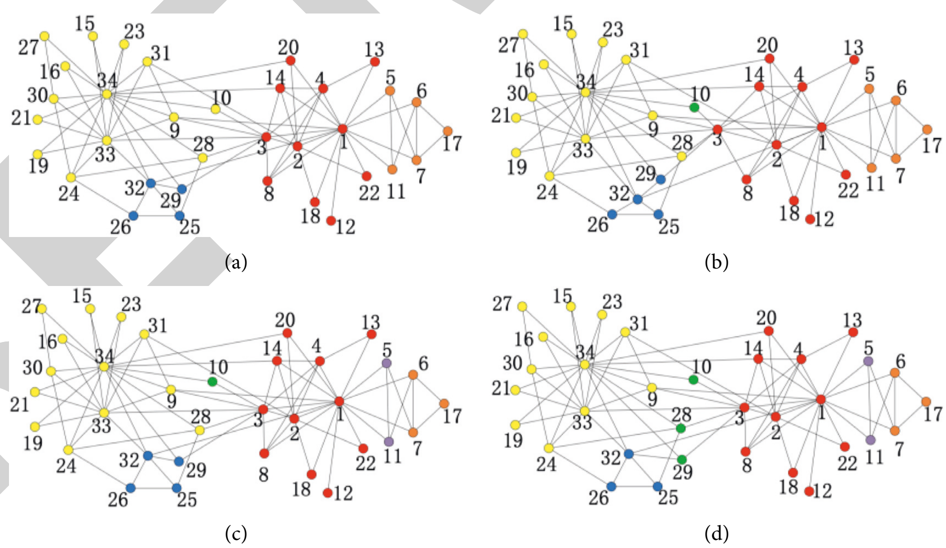


FIGURE 3: A tiny complex network with local weak edges.



(a) (b)

FIGURE 4: A typical part of complex network with two communities. (a) Community 1. (b) Community 2.



(a) (b)

(c) (d)

FIGURE 5: A tiny complex network with a number of nodes.

in the figure, we can know that the algorithm divides the network graph into 15 associations and the nodes belonging to the same association are labeled with the same color. The edges within each association are relatively tightly connected, while the edges between associations are sparsely connected. Furthermore, the right part of Figure 8 shows a tree diagram of the association splitting process in the risk network, and the left part shows the change in the similarity-based modularity index value based on the formula as each association splitting behavior occurs. The tree diagram on the right side of the figure shows that the algorithm proposed in this section can effectively discover the hierarchical

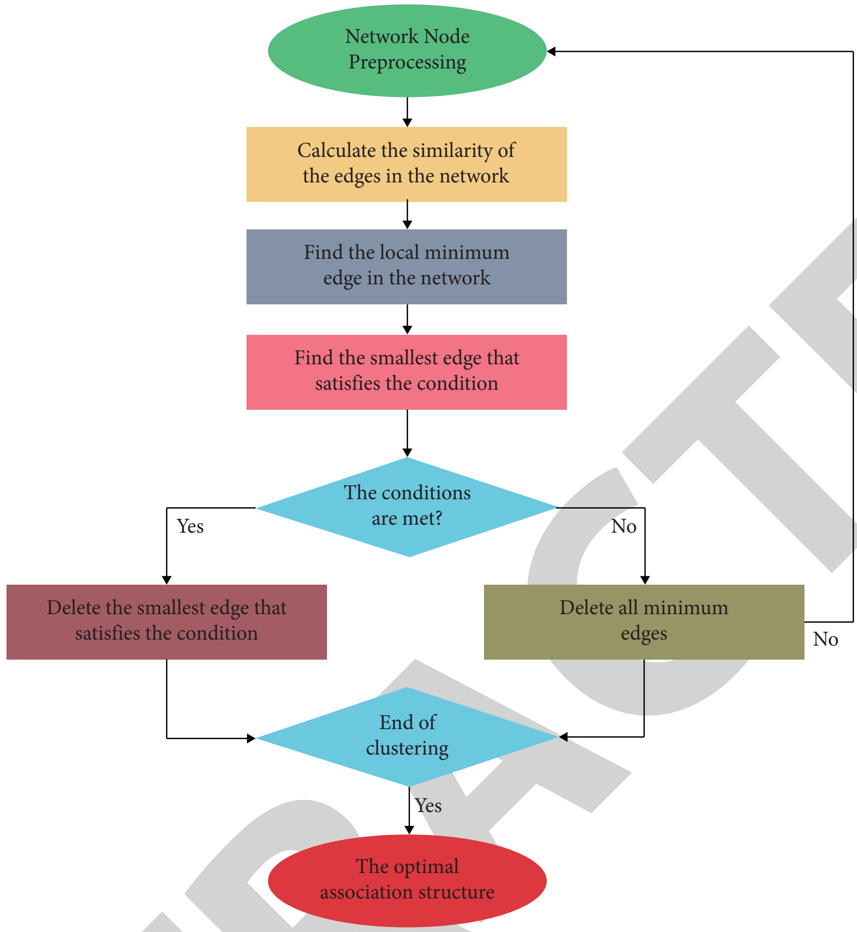


Figure 6: The workflow of the tiny complex network algorithm.

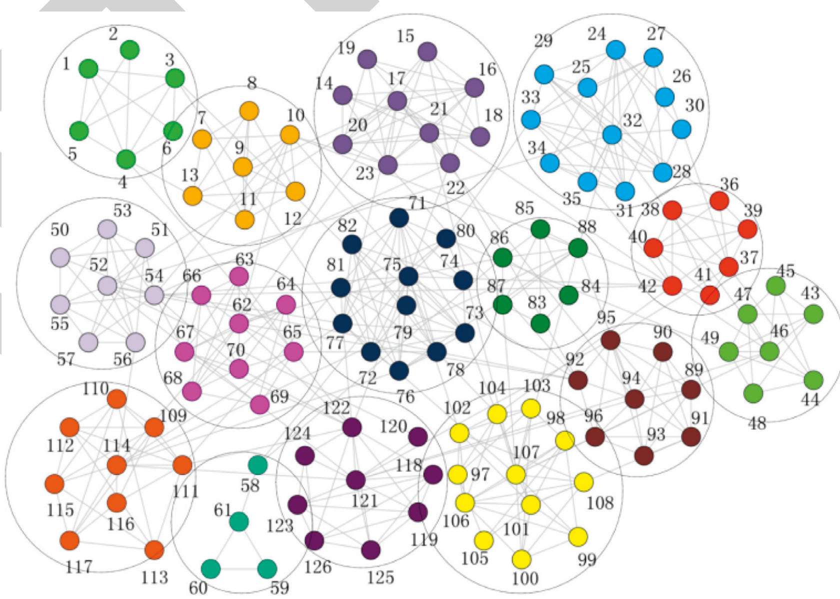


Figure 7: Community split tree diagram.

association structure. For example, association I is the first one to be discovered, which is a complete graph with three nodes and has only one edge between node 61 and the rest of the network, so the association structure is isolated. After that, it is gradually split into several associations according to the criterion of similarity between agricultural economic

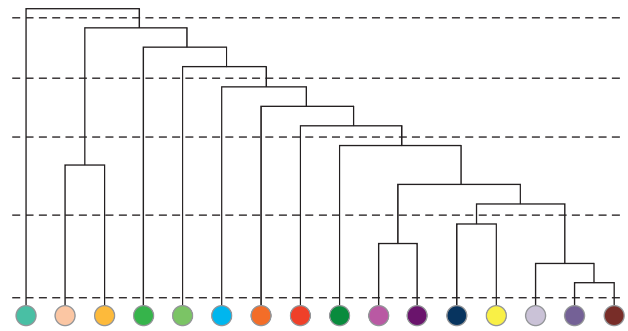


Figure 8: Community split tree diagram.

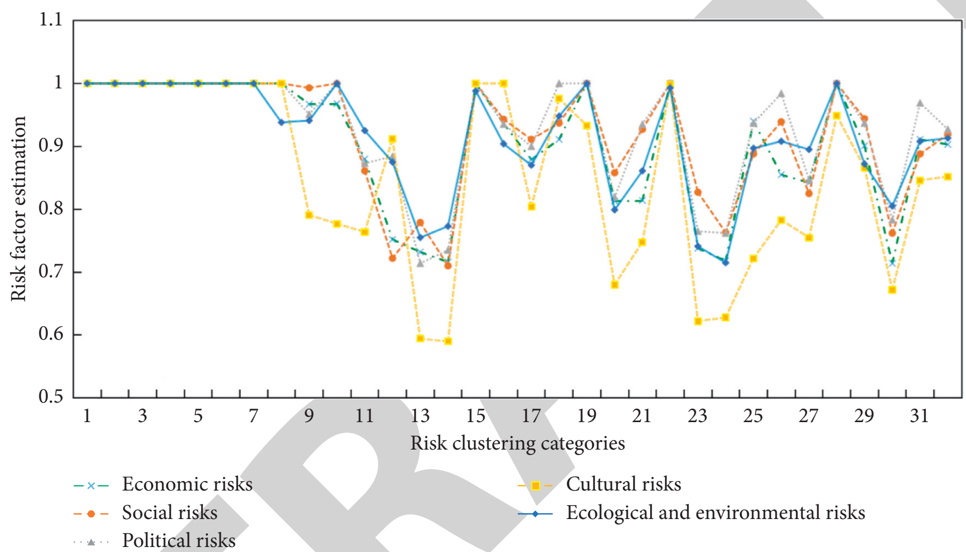


Figure 9: Community split tree diagram.

management risks. Each time the new association structure is split, the modularity index based on similarity increases, thus demonstrating the gradual rationalization of the association classification results.

The statistical yearbook of the National Bureau of Statistics, statistical bulletin, and China Rural Statistical Yearbook were used to model each risk dimension. Figure 9 shows the clustering analysis and coefficient estimation for each risk dimension. The results show that the risk dimensions tend to be consistent in terms of trend changes and the clustering pattern is also relatively consistent despite the slight differences, indicating that the algorithm has high applicability in assessing risk.

5. Conclusions

Current agricultural economic management focuses on the evaluation of economic and ecological risks while neglecting other possible risk dimensions, making it difficult to adapt to the changing needs of the market. More importantly, the current research still adopts traditional analysis methods and greatly ignores the interrelationships among risk dimensions. Therefore, this article first proposes five risk dimensions of agricultural economic management—economic, social, political, cultural, and ecological and environmental risks—and further clarifies the logical network relationships of different risk dimensions based on the identification of risk evaluation dimensions of agricultural economic management. The complex network analysis is used to further identify and evaluate the risks of agricultural economic management. Therefore, this article adopts a factor analysis technique based on complex network theory and empirically tests the applicability of this multivariate statistical calculation method. This article shows that the risks in agricultural and forestry economic management is of a complex network structure and the assessment of risks cannot be cut simply from the evaluation of each dimension; however, the intrinsic network structure of risks must be considered comprehensively to produce correct estimation results. This article theoretically innovates the identification of risk dimensions of agricultural economic management and applies factor analysis techniques based on complex network theory methodologically, but there are still the following shortcomings: first, this article has not yet developed an ephemeral trend analysis, and expanding the analysis on the time scale can further resolve the volatility of risks. Second, the method adopted in this article still lacks the support of a large amount of data and the cross-sectional comparison with other evaluation methods. Finally, there is still room for improvement in the way the risk dimensions are classified in this article.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this article.

References

[1] D. Liu, X. Zhu, and Y. Wang, "China's agricultural green total factor productivity based on carbon emission: an analysis of evolution trend and influencing factors," *Journal of Cleaner Production*, vol. 278, Article ID 123692, 2021.

[2] T. Kenderdine, "Insurance plus futures: agricultural commodity price reform in China," *Asia & the Pacific Policy Studies*, vol. 5, no. 2, pp. 331–346, 2018.

[3] S. J. H. Shahzad, J. A. Hernandez, K. H. Al-Yahyaee, and R. Jammazi, "Asymmetric risk spillovers between oil and agricultural commodities," *Energy Policy*, vol. 118, pp. 182–198, 2018.

[4] C. M. Boyd and M. F. Bellemare, "The microeconomics of agricultural price risk," *Annual Review of Resource Economics*, vol. 12, no. 1, pp. 149–169, 2020.

[5] A. Burliai, O. Burliai, A. Revutska, L. Smolii, and L. Klymenko, "ОрҐанізаційно-економічні ризики еколоҐізації сільськоҐо ҐосЦодарства," *Agricultural and Resource Economics: International Scientific E-Journal*, vol. 7, pp. 96–114, 2021.

[6] M. Graubner, K. Salhofer, and C. Tribl, "A line in space: pricing, location, and market power in agricultural product markets," *Annual Review of Resource Economics*, vol. 13, no. 1, pp. 85–107, 2021.

[7] B. Gong, "Interstate competition in agriculture: cheer or fear? Evidence from the United States and China," *Food Policy*, vol. 81, pp. 37–47, 2018.

[8] K. Purnhagen, "The increasing overlap of agricultural, free movement and competition law in the EU," *European Law Review*, vol. 46, no. 1, pp. 20–36, 2021.

[9] H. Do and T. Kompas, "Economic efficiency versus social equity: the productivity challenge for rice production in a "greying"rural Vietnam," *World Development*, vol. 148, Article ID 105658, 2021.

[10] E. Augeraud-Véron, G. Fabbri, and K. Schubert, "Volatility-reducing biodiversity conservation under strategic interactions," *Ecological Economics*, vol. 190, Article ID 107193, 2021.

[11] A.-F. M. El-Sayed, A. M. Nasr-Allah, M. Dickson, and C. Gilmour, "Analysis of aquafeed sector competitiveness in Egypt," *Aquaculture*, vol. 547, Article ID 737486, 2022.

[12] S. Bernstein, E. Colonnelli, D. Malacrino, and T. McQuade, "Who creates new firms when local opportunities arise?" *Journal of Financial Economics*, vol. 143, no. 1, pp. 107–130, 2022.

[13] J. Chen, "Multidimensional analysis model of agricultural product supply chain competition based on mean fuzzy," *Journal of Intelligent and Fuzzy Systems*, vol. 41, pp. 1–12, 2021.

[14] J.-P. Chavas, S. Di Falco, F. Adinolfi, and F. Capitanio, "Weather effects and their long-term impact on the distribution of agricultural yields: evidence from Italy," *European Review of Agricultural Economics*, vol. 46, no. 1, pp. 29–51, 2019.

[15] M. Roznik, C. Brock Porth, L. Porth, M. Boyd, and K. Roznik, "Improving agricultural microinsurance by applying universal kriging and generalised additive models for interpolation of mean daily temperature," *The Geneva Papers on Risk and Insurance - Issues and Practice*, vol. 44, no. 3, pp. 446–480, 2019.

[16] Y. Sheng, S. Zhao, and S. Yang, "Weather shocks, adaptation and agricultural TFP: a cross-region comparison of Australian Broadacre farms," *Energy Economics*, vol. 101, Article ID 105417, 2021.

[17] P. M. Regan, H. Kim, and E. Maiden, "Climate change, adaptation, and agricultural output," *Regional Environmental Change*, vol. 19, no. 1, pp. 113–123, 2019.

[18] S. Sun and Q. Chen, "Household responsibility system and China's agricultural Growth revisited: addressing endogenous institutional change," *Economics of Transition and Institutional Change*, vol. 28, no. 4, pp. 537–558, 2020.

[19] Y. Tang, H. Cai, and R. Liu, "Farmers' demand for informal risk management strategy and weather index insurance: evidence from China," *International Journal of Disaster Risk Science*, vol. 12, no. 2, pp. 281–297, 2021.

[20] J. Zhang, Z. Zhang, and F. Tao, "Rainfall-related weather indices for three main crops in China," *International Journal of Disaster Risk Science*, vol. 11, no. 4, pp. 466–483, 2020.

[21] X.-B. Jin, X.-H. Yu, X.-Y. Wang, Y.-T. Bai, T.-L. Su, and J.-L. Kong, "Deep learning predictor for sustainable precision agriculture based on internet of things system," *Sustainability*, vol. 12, no. 4, p. 1433, 2020.

[22] J. Guo, J. Jin, Y. Tang, and X. Wu, "Design of temperature insurance index and risk zonation for single-season rice in response to high-temperature and low-temperature damage: a case study of jiangsu province, China," *International Journal of Environmental Research and Public Health*, vol. 16, no. 7, p. 1187, 2019.

[23] Z. Shi, H. Huang, Y. Wu, Y.-H. Chiu, and S. Qin, "Climate change impacts on agricultural production and crop disaster area in China," *International Journal of Environmental Research and Public Health*, vol. 17, no. 13, p. 4792, 2020.

[24] L. Minale, "Agricultural productivity shocks, labour reallocation and rural–urban migration in China," *Journal of Economic Geography*, vol. 18, no. 4, pp. 795–821, 2018.

[25] S. Chonabayashi, T. Jithitikulchai, and Y. Qu, "Does agricultural diversification build economic resilience to drought and flood? Evidence from poor households in Zambia," *African Journal of Agricultural and Resource Economics*, vol. 15, pp. 65–80, 2020.

[26] B. Parisse, A. Pontrandolfi, C. Epifani, R. Alilla, and F. De Natale, "An agrometeorological analysis of weather extremes supporting decisions for the agricultural policies in Italy," *Italian Journal of Agrometeorology*, no. 1, pp. 15–30, 2020.

[27] S. Das and K. Goswami, "Progress in agricultural vulnerability and risk research in India: a systematic review," *Regional Environmental Change*, vol. 21, no. 1, pp. 1–18, 2021.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9843864, 1 page
https://doi.org/10.1155/2023/9843864

Retraction

Retracted: Research on Enterprise Financial Management and Prediction System Based on SaaS Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Zhang and F. Zhou, "Research on Enterprise Financial Management and Prediction System Based on SaaS Model," *Security and Communication Networks*, vol. 2022, Article ID 3218903, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3218903, 9 pages
https://doi.org/10.1155/2022/3218903

Research Article

Research on Enterprise Financial Management and Prediction System Based on SaaS Model

Qianying Zhang[1] and Fang Zhou[2]

[1]*School of Economics and Management, Hebei Oriental University, Langfang 065000, China*
[2]*College of Artificial Intelligence, Hebei Oriental University, Langfang 065000, China*

Correspondence should be addressed to Fang Zhou; zhoufang@hou.edu.cn

Received 21 February 2022; Revised 5 March 2022; Accepted 14 March 2022; Published 13 April 2022

Academic Editor: Chin-Ling Chen

In order to supervise and forewarn the sustainable operation ability of enterprises efficiently and accurately, this paper proposes an enterprise financial management and forecasting system based on SaaS model. First of all, in order to continue to effectively predict and analyze the enterprise finance, first analyze and extract the report data in the financial system. Then, by building a deep belief network model to predict the enterprise financial data, in order to reduce the cost of enterprises, the financial system designed in this paper chooses the cloud technology service framework based on SaaS model. Finally, in order to analyze the risk identification performance of the financial management and prediction system in this paper, the risk sample data of an enterprise's financial system is selected for simulation test. The results show that the correct rate of risk identification of the financial management system designed in this paper is higher than other comparison systems, which speeds up the speed of risk identification of the financial information management system, and has certain practical application value.

1. Introduction

With the continuous development of economy, financial management is widely used in enterprises. Due to the increasing probability of economic crisis, the probability of bankruptcy of enterprises is getting higher and higher. Therefore, improving financial management level can reduce the probability of bankruptcy of enterprises [1]. Financial risk identification is an important branch of financial management. Financial risk identification can help enterprises develop certain risk control measures and control financial risk identification within a certain range. It can better promote the development of enterprises, so financial risk identification has always been regarded as a major research topic [2].

In the past few decades, domestic and overseas scholars have conducted extensive research on financial risk identification. At first, people used the theory of statistics to conduct financial risk identification research [3], such as the financial risk identification method based on decision tree, the financial risk identification method based on random forest algorithm, the financial risk identification method based on XGBoost algorithm [4], and the financial risk identification method based on multivariate discriminant analysis. They assume that financial risk is a fixed law of change, such as linearity and periodicity. However, there are many factors involved in financial risk, which make the change of financial risk complicated and time-varying. Traditional statistical theory is difficult to accurately grasp the characteristics of financial risk changes, resulting in low accuracy of financial risk identification [5]. Soon afterwards there are financial risk identification methods based on modern statistical theory, such as BP neural network financial risk identification method [6] and RBF neural network financial risk identification method which support vector machine financial risk identification method [7]. Their financial risk identification effect is obviously better than traditional statistical theory. In the process of financial risk identification modelling [8], the neural network requires a large number of samples, while the financial risk samples are usually small, which makes difficult to meet the requirements of large samples. As a result, the financial risk identification results of neural network often show overfitting phenomenon, and the credibility of financial risk identification results is low. Although support vector machine will not produce "overfitting" financial risk identification results, the low modelling efficiency leads to a long time of financial risk identification [9].

In recent years, with the continuous development of data mining technology, deep belief network has been greatly developed, which provides a new research tool for risk identification modelling of financial information management system [10]. Because of the complexity of risk change in financial information management system, in order to improve the accuracy of risk identification in financial information management system, this paper proposes a prediction model based on improved deep belief network. Two-dimensional rotation crossing (TRC) strategy and adaptive mechanism are introduced to improve the accuracy of model prediction.

This paper mainly has the following innovations:

(1) The report data in the financial system is analyzed and extracted
(2) The parameters of hidden layer are optimized by deep belief network combined with improved differential evolution algorithm
(3) The financial system designed in this paper chooses the cloud technology service framework based on SaaS model to reduce the cost of enterprises

This paper consists of five main parts: the first part is the introduction, the second part is prediction model based on improved deep belief network, the third part is financial management system based on SaaS model, the fourth part is the experiments, and the fifth part is the conclusion; besides there are abstracts and references.

2. Prediction Model Based on Improved Deep Belief Network

2.1. Data Preprocessing. In order to continue to forecast and analyze effectively, it is necessary to analyze and extract various forms and statements data from the financial system of the enterprise [11]. Data cleaning is used to generate comma-separated values (CSV) data as required.

2.1.1. Data Cleaning. Aiming at the problem of low integrity and much overlap of expenditure funds in corporate financial institutions, the data cleaning adopted in this paper is divided into four steps.

(1) *Missing value cleaning*: Set the threshold value of miss judgment to 80%, and select the original data according to the standard, remove the feature columns exceeding the threshold value, and fill the missing value in the area with "0" value
(2) *Format content cleaning*: Set the imported data to a unified storage format, for example, 2019-02-21
(3) *Repeated content cleaning:* Then, the data is filtered again, and multiple feature columns with high content repetition are deleted to keep only one of them, which is conducive to dimensionality reduction
(4) *Nonrequired data cleaning*: The irrelevant data that are not in the predicted time span are deleted, and only the sample data with the minimum time span of 1 month are retained

After the above four steps, all the processed data is saved in the required CSV format

2.1.2. Feature Selection. The characteristics of each column in the data sample need to be reasonably selected, so as to reflect the required relational mapping and avoid overfitting as much as possible and to strengthen the ability of multiple model generalization.

In this paper, L1 norm regularization method is used for feature selection, which can be effectively applied to nonlinear scenes [12]. The L1 norm scores of all statistical features were calculated. 0.6 was set as the selection threshold in this paper, and features with scores less than 0.6 were deleted to complete the feature selection process.

2.1.3. Normalization. After cleaning the financial data, it is also necessary to unify the value range of the actual sample values in order to unify the scale of the sample characteristics. This paper adopts mean variance normalization to process all data samples, which are uniformly expressed as numbers between [0, 1], as shown in the following equation:

$$I_{\text{scale}} = \frac{i - \min}{\max - \min}, \tag{1}$$

where min represents the minimum eigenvalue and max represents the maximum eigenvalue.

2.1.4. Generation of Sliding Samples. Since financial forecasting is a periodic work, time span needs to be set, similar to the window frame in the graphic image processing mechanism. In this paper, 2 years is set as the time span for sliding selection of data samples. Too short or too long time span will have a certain impact on the performance of the prediction, and 2 years is the empirical value of multiple experiments.

2.2. Deep Belief Networks. DBN is a neural network classification, recognition, and prediction model with multiple hidden layers. Compared with shallow machine learning and traditional neural network, DBN with a large number of hidden layers has good feature extraction and recognition ability for abnormal data flow of enterprise finance. Through multilayer nonlinear transformation, deep abstract features are trained from complex enterprise financial data, and the internal correlation of data is described. The layer-by-layer training method overcomes the defect that feedforward neural network is easy to fall into local minimum value and obtains better prediction results and faster convergence speed.

2.2.1. Sample Input. This paper adopts Support Vector Classification (SVC) model to collect four kinds of historical sample data sets and obtains the current enterprise financial situation assessment grade set $V = \{v_x\}$ according to the evaluation method of existing literature.

v_x is the situation assessment grade value of enterprise finance at moment x, and then the assessment set V is converted into the input data set N of DBN network. As shown in formula (2) i_a is the input feature vector, j_a is the label value, a is the serial number of the sliding window, t is the size of the sliding window, the sliding step of the window is 1, $t \le T$, $a \le T - t$.

$$N = \{i_a, j_a\} \Rightarrow \{(li_a, \ldots, li_{a+t-2}), (lj_a, \ldots, lj_{a+t-1})\}. \quad (2)$$

2.2.2. Training Phase. DBN uses two steps of unsupervised pretraining and supervised global fine-tuning to adjust the weight of neural network. In the pretraining, Restricted Boltzmann machine (RBM) was trained separately, layer by layer from low level to high level. In fine-tuning stage, BP neural network is used to fine-tune the weight and bias of DBN. The DBN structure is shown in Figure 1.

RBM obtains the weights of the generated model through pretraining in an unsupervised, layer-by-layer greedy manner. During the neural network training, the visual value is mapped to the hidden layer node, and the hidden layer node is reconstructed as the visible node. The value of each node is in the set {0, 1}; that is, there exists any x, y such that $q_x \in \{0, 1\}$, $b_y \in \{0, 1\}$, $m_{x,y}$ are the weights between the visible node and the hidden node, and the offset of the visible node $c = (c_1, c_2, \ldots, c_x)$, and the offset of hidden layer node $h = (h_1, h_2, \ldots, h_y)$. For RBM with x visible nodes and y hidden layer nodes, b and q represent the states of hidden layer nodes and visible nodes, respectively. Set a set of states (b, q), and the energy function of RBM is defined as the following formula:

$$E(q, b|\phi) = -\sum c_x q_x - \sum h_y b_y - \sum\sum m_{xy} q_x b_y, \quad (3)$$

where $\phi = (c_x, h_y, m_{xy})$ represents the joint distribution law of (b, q) obtained by formula (3) for RBM parameters, as shown in the following formula:

$$u(q, b|\phi) = F(\phi)^{-1} \exp[q, b], \quad (4)$$

where E is the expected value and $F(\phi)$ is the normalized factor, as shown in the following formula:

$$F(\phi) = \sum\sum \exp(-E(q, b|\phi)). \quad (5)$$

When the visible layer is known, all nodes of the hidden layer are independent of each other, and the probability distribution of the y-th node in the hidden layer is shown in the following formula:

$$\left\{ u(Q|B) = \prod_x u(b_y|q), u(b_y = 1|q) = f\left(h_x + \sum m_{xy} q_x\right), u(b_y = 0|q) = 1 - u(b_y = 1|q), \right. \quad (6)$$

where u is the probability and f is the sigmoid activation function. Similarly, if the hidden layer is known, the probability distribution of the xth node in the visible layer is obtained. Because the data in the sample training set has the label of enterprise financial situation data source, in the RBM training at the top level, in addition to the dominant neurons, there should also be neurons representing classification labels in the visible layer. Here, for each group of training data, the corresponding label neurons are opened as 1, while the others are closed and set as 0. The algorithm of DBN training process is obtained.

2.3. Prediction Model. Because the layer number of hidden layer unit of DBN network is mostly set by experience, and the related parameters of hidden layer are sensitive to this, improper selection may lead to serious decline in accuracy of prediction results or too long training time. In order to solve these problems, Differential Evolution (DE) algorithm is integrated into deep belief network to simplify network structure and build a good deep learning model [13]. The core problem of the enterprise financial situation prediction model based on IDE-DBN is to determine the number of hidden node layers and the weight and bias between layers. The improved DE algorithm enhanced the global search performance and further reduced the local extremum [14]. The basic idea is to map the parameters of hidden layer nodes of DBN network used for enterprise financial situation prediction to the initial target individuals of differential vector space evolution [15]. By judging the fitness level of the new individuals and the old individuals, the weak ones are selected and the strong ones are retained. After several iterations, the target search is guided to the optimal solution with low error. The mathematical description of DE algorithm is shown in the following formula:

$$\mathrm{DE} = \{a_0, w_0, l_0, v_0, f_0, h_0, c_0\}, \quad (7)$$

where a_0 is the number of nodes in hidden layer of DBN network, w_0 is the parameter of hidden layer of DBN, l_0 is the population size, v_0 is the fitness function of individual, and three operations are defined: f_0 is the replication operation, h_0 is the crossover operation, and c_0 is the mutation operation.

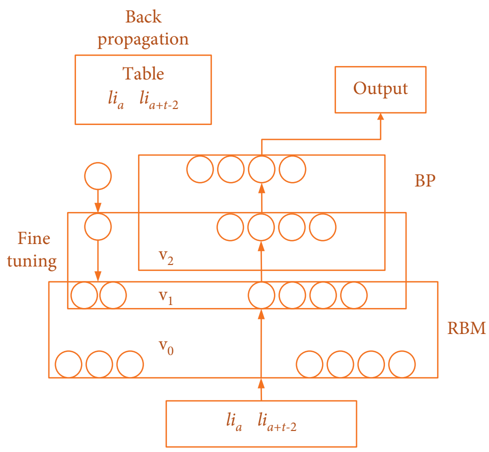


FIGURE 1: Structure of DBN.

2.3.1. Standard DE Algorithm. The main process of standard DE algorithm is shown in Figure 2.

In the standard DE algorithm, the difference vector between the parent individuals is generated, and it is superimposed on a random target individual. The optimal individual is retained and added to the next generation population through the crossover and selection operation between the new offspring and the parent. Algorithm optimization is a process of constant transformation in search space. However, it is easy to fall into local optimality, so the crossover process is improved in this paper. Population initialization: Set TU to represent population size and D to represent individual dimension and randomly generate the initial population in the problem decision space according to the following formula:

$$i_{x,0} = \left(i_{r1,0}, i_{r2,0}, \ldots, i_{rD,0}\right) \quad x = 1, 2, \ldots, \mathrm{TU}, \tag{8}$$

where $i_{x,0}$ is the ith individual of the population of the 0th generation, and the individual is valued in each dimension according to the following formula:

$$i_{xy,0} = L_y + \mathrm{rand}\left(B_y - L_y\right), \quad y = 1, 2, \ldots, D, \tag{9}$$

where (B_y, L_y) represents the value range of the yth dimension, $i_{xy,0}$ represents the yth dimension component of the xth individual of the initial generation, and RAND represents the random number evenly distributed on the interval (0, 1).

Mutation means that, after initialization, DE algorithm generates a population composed of NP experimental vectors through mutation and recombination operation of the population and produces mutant individuals by perturbation of target individuals. Common mutation strategies are as follows: the DE/RAND/1 differential mutation operation adds a scalable and randomly selected vector to a third-party vector, ss shown in the following formula:

$$v_{i,G} = x_{r1,G} + F \cdot \left(x_{r2,G} - x_{r3,G}\right). \tag{10}$$

In DE/ Best/1 strategy, the differential variation operation with a scalable and randomly selected vector is added to the vector with the best performance of the basis vector in the current population, as shown in the following formula:

$$q_{x,A} = i_{\mathrm{best},A} + F \cdot \left(i_{r1,A} - i_{r2,A}\right). \tag{11}$$

In DE/current-to-best/1 strategy, the differential mutation operation and two randomly selected scalable vectors are added to the third-party vector, and the vector difference contains the vector with the best performance of the basis vector in the current population, as shown in the following formula:

$$q_{x,A} = i_{x,A} + F \cdot \left(i_{\mathrm{best},A} - i_{x,A}\right) + F \cdot \left(i_{r1,A} - i_{r2,A}\right). \tag{12}$$

Crossover: the experimental individual is obtained by crossover operation between the parent individual and the mutant individual.

$$p_{xy,A} = \begin{cases} q_{xy,A}, & \text{if } \mathrm{rand}\,(0,1) \leq \mathrm{CR} \text{ or } y = y_{\mathrm{rand}}, \\ i_{xy,A}, & \text{otherwise.} \end{cases} \tag{13}$$

Selection refers to the selection of experimental individuals and paternal individuals based on the principle of greed, and the excellent individuals will enter the next generation.

2.3.2. Adaptive Mechanism. The difference vector is the disturbance of the variable-dimensional decision variables of i_{rx} and A. And F controls the degree of disturbance. If the generated difference vectors are relatively close apart in the search space, F should take a larger value; otherwise the perturbation quantity is too small to be conducive to global search in the early stage of evolution. If the generated difference vectors are far apart in the search space, F should take a smaller value; otherwise the disturbance is too large, which is not conducive to local search in the early stage of evolution. The value of F should be adjusted adaptively according to the relative position of the two individual vectors generating difference vectors in space to balance the contradiction between local search and global search.

$$F_{x,A+1} = \begin{cases} F_{\min} + \mathrm{rand}\left(F_{\max} - F_{\min}\right) \text{ if } f\left(p_{x,A}\right) > f\left(i_{x,A}\right) \\ F_{x,A}, \text{ otherwise} \end{cases}, \tag{14}$$

where $F_{x,A+1}$ represents the scaling factor of the xth individual of the $A+1$ generation, $F_{x,A}$ represents the scaling factor of the ith individual of the A generation, and an appropriate scaling factor is conducive to the generation of offspring with high survival rate. Therefore, a new Cr_i can be generated according to formula (15) by recording the factor parameters of the most recent history.

$$\mathrm{Cr}_{x,A+1} = \begin{cases} \mathrm{rand}\, t_x(v, \varepsilon) \text{ if } p_{x,A}) > f\left(i_{x,A}\right) \\ \mathrm{CR}_{x,A} \text{ otherwise} \end{cases}, \tag{15}$$

where Cr_x follows a Gaussian distribution and represents the crossover rate of the xth individual in the A generation, while $\mathrm{rand}\, t_x(v, \varepsilon)$ is the Gaussian distribution of mean v and standard deviation ε.

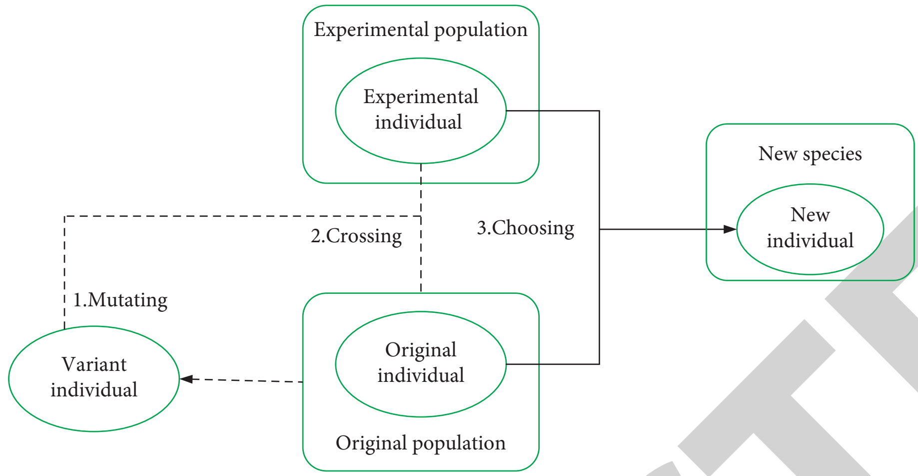


Figure 2: Standard DE algorithm flow.

2.3.3. Two-Dimensional Rotation Crossover Algorithm. With the deepening of evolutionary algebra, the standard DE algorithm cannot cope with the adaptive changes of population diversity and the global optimal solution cannot guarantee convergence in the feasible solution space. By introducing a two-dimensional rotational crossover algorithm, the offspring of the target and the mutant are generated around the target. Suppose that the rotation radius of the population containing NP d-dimension individuals is shown in the following formula:

$$\begin{cases} \vec{r}_{x,A} = \theta\left(\vec{q}_{x,A} - \vec{i}_{x,A}\right), \\ \vec{R}_{x,A} = \dfrac{1}{2}\vec{r}_{x,A}\vec{R}_{x}. \end{cases} \tag{16}$$

The direction control parameter θ is set to ensure that the offspring can be evenly distributed near the mutant and the target. When $\theta = 1$, the direction of the cross vector is from $\vec{q}_{x,A}$ to $\vec{i}_{x,A}$, and when $\theta = -1$, the vector is reversed. $|\vec{R}_{x,A}|$ is the distance between the target individual and the target individual, $\vec{R}_{x,A}$ is the radius of rotation distributed near the new offspring, ω is the vector $\vec{R}_{x,A}$ rotation angle of A, A is the iteration number, $\vec{R}$ is the vector controlling rotation, and its value is shown in the following formula:

$$\vec{R} = C_1 (A+1)^{-1/2} R'. \tag{17}$$

C_1 is a universal constant, and the modulus of $\vec{R}$ decreases with the increase of A, which is conducive to increasing the search accuracy and accelerating the convergence of TRC-DE algorithm. The adjusted control vector factor R' follows the Cauchy distribution as shown in the following formula:

$$R' = \text{rand Cauchy}. \tag{18}$$

R' is a matrix of $NP \times D$, whose individuals satisfy the Cauchy distribution. Because the Cauchy distribution has a long tail, the selection range of progeny produced by controlling the Cauchy distribution increases.

According to the above analysis, with parent $\vec{i}_{x,A}$ and variant $\vec{q}_{x,A}$ as the center, and with rotation control vector $\vec{R}$ and direction control vector θ acting on the mode of $\Delta R = \theta \cdot R \cdot r$ as the rotation radius, the cross vector $\vec{K}_{x,A}$ is obtained. The rotation crossover operation is shown in the following equation:

$$\vec{K}_{x,A} = \begin{cases} \theta\vec{R}_{x,A}\left(\vec{q}_{x,A} - \vec{i}_{x,A}\right) + \vec{q}_{x,A} & \text{rand}_{xy} \le \text{Cr or } y = y_{\text{rand}}, \\ \theta\vec{R}_{x,A}\left(\vec{q}_{x,A} - \vec{i}_{x,A}\right) + \vec{i}_{x,A} & \text{otherwise}. \end{cases} \tag{19}$$

The diversity of progeny can be increased by incorporating two-dimensional rotational crossover.

In this paper, an example of rotating vectors in two-dimensional space is given to prove how to improve the population diversity and convergence of standard DE algorithm by introducing two-dimensional rotating crossover algorithm. Assuming $\vec{R}_x = (r_1, r_2)(r_1 \leqslant r_2)$, $\vec{i}_{x,A} = (i_1, i_2)$, $\vec{q}_{x,A} = (q_1, q_2)$, $\theta = 1$, substitute into the following formula:

$$\begin{cases} \dfrac{\left(\vec{q}_{x,A} - \vec{i}_{x,A}\right)}{2} = (\Delta i_1, \Delta i_2) \\ \vec{R}_{x,A} = \left(\vec{q}_{x,A} - \vec{i}_{x,A}\right)\vec{R}_x \end{cases}, \tag{20}$$

where $\Delta i_1 = q_1 - i_1$, $\Delta i_2 = q_2 - i_2$, the rotation vector can be expressed as $\vec{R}_{x,A} = (r_1 \Delta i_1, r_2 \Delta i_2)$, and the modulus of the rotation vector can be expressed as

$$\begin{aligned} \left|\vec{R}_{x,A}\right| &= \sqrt{\left(r_1 \Delta i_1\right)^2 + \left(r_2 \Delta i_2\right)^2)} \\ &= r_2 \sqrt{\left(r_1/r_2 \Delta i_1\right)^2 + \left(\Delta i_2\right)^2)}. \end{aligned} \tag{21}$$

Let $r_1 \le r_2$, then $r_1 \le |\vec{R}_{x,A}| \le r_2$, the modulus $|\vec{R}_{x,A}|$ of $\vec{R}_{x,A}$ can vary in size with different coordinate values of $\vec{R}_x$, and the rotation angle ω of $\vec{R}_{x,A}$ can be obtained by the following formula:

$$\omega\left(\vec{R}_{x,A}\right) = \text{grcngt}\left(\frac{r_1}{r_2}\cdot\frac{\Delta i_1}{\Delta i_2}\right). \tag{22}$$

In formula (22), there are three possible situations: if $r_1 = r_2$, the vector $\vec{R}_{x,A}$ has the same direction as $\vec{q}_{x,A} - \vec{i}_{x,A}$. If $r_1 < r_2$, then ω will have a greater angle than the vectors $\vec{q}_{x,A} - \vec{i}_{x,A}$. If $r_1 > r_2$ then the angle ω will be smaller than the vectors $\vec{q}_{x,A} - \vec{i}_{x,A}$, and the magnitude and ω of the rotation vector vary with the coordinates of the vector $\vec{R}_i$.

2.3.4. IDE-DBN Algorithm. IDE-DBN is a combination of TRC-DE algorithm and DBN algorithm. It treats DBN weights and bias values as populations and uses them to perform mutation, crossover, and selection operations.

3. Financial Management System Based on SaaS Model

In the use of cloud computing, there are three service models. In order to ensure that the quoted mode is suitable for the financial risk control system, it is necessary to compare the three service modes (as shown in Table 1). According to the comparison results, select the mode suitable for this system (Figure 3).

Through the comparison of service modes, SaaS service is selected as the service mode of the cloud computing method. This method can effectively reduce the cost of enterprises and reduce the threshold and risk of enterprise informatization. The implementation of cloud computing method is set as three parts: cloud collection of financial data, cloud processing, and financial abnormal data.

In the cloud collection part, big data network is used to complete the acquisition of financial information. The financial information is processed by distributed computing method to obtain abnormal data in the financial data. At this point, the cloud computing part of the design is complete.

4. Experiments

4.1. Test Data Set. In order to analyze the risk identification performance of data mining financial information management system, a risk sample data of financial information management system is selected for simulation test, and the sample data is shown in Figure 4.

As it can be seen from Figure 4, the risk changes of the information management system are complicated. It has a variety of characteristics, such as regularity, time variability, randomness, and so on. In order to make the data mining financial information management system risk identification results more convincing, the traditional financial information management system risk identification is a comparative test. They are the financial information management system risk identification method based on [16] and the financial information management system risk identification method based on [17]. The correct rate of risk identification, rejection rate, and modelling time of risk identification of financial information management system are selected as evaluation indicators.

Table 1: Comparison of cloud computing service modes.

Mode code	Name	Service content
1	SaaS service 7	Software, data, and information
2	PAAS services	Operation, support, and development
3	LAAS service 8	Server, database, and storage

4.2. Division of Training Samples and Test Samples. In the process of risk identification modelling of financial information management system, training is the first step. Therefore the selection of training samples is very critical. In order to reflect the fairness of the experimental results, each method carries out five simulation experiments of risk identification of financial information management system and adopts different numbers of training samples to train the model. The test samples are mainly used to test the generalization ability of the risk identification method of the financial information management system. The partitioning results of the training samples and test samples of each experiment are shown in Table 2.

4.3. Results and Analysis. The correct number identified by the three methods from the identification training samples of the financial information management system is counted, and the risk identification accuracy can be obtained by comparing with the number of test samples. The modelling time is the identification training and testing time of the financial information management system, as shown in Figures 5 and 6.

Analysis of them leads to the following conclusions.

(1) Among all the methods, [16] has the lowest risk identification effect of financial information management system. This is mainly because [16] method is a simple linear modelling technology, which cannot fully describe the rule of system risk change, and the risk identification results of financial information management system are not ideal.

(2) The risk identification effect of financial information management system based on [17] is better than that of [16]. The risk identification error of financial information management system based on [16] is reduced. However, [17] also has defects, such as unstable risk identification results and poor reliability of financial information management system.

(3) The risk identification effect of financial information management system based on data mining is better than that of [16] and [17], which reduces the risk identification error of financial information management system and greatly reduces the risk rejection rate of financial information management system. The results of risk identification of financial information management system are more reliable, which

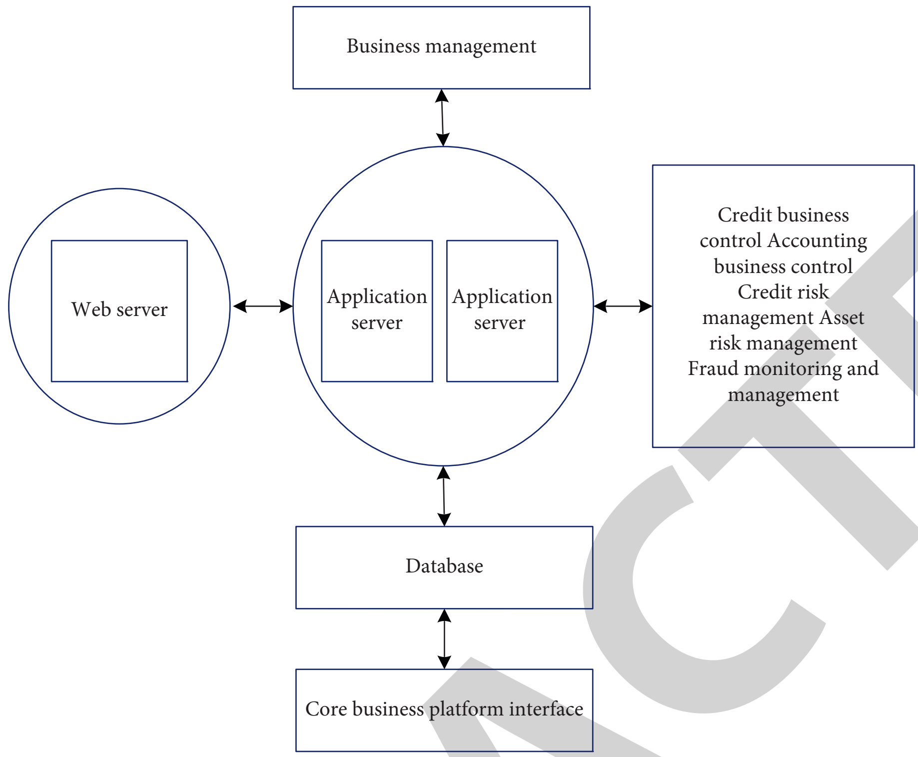


Figure 3: System software module diagram.

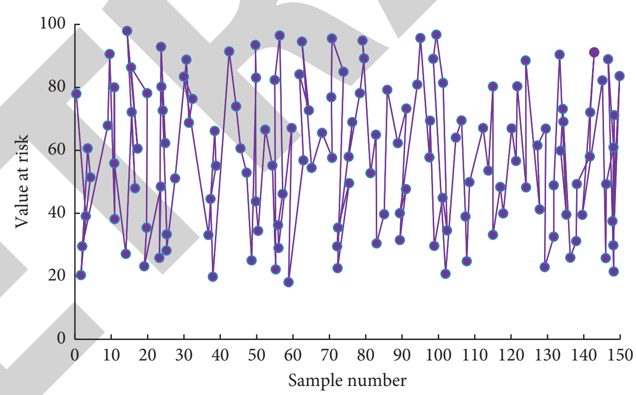


Figure 4: Experimental sample data set used in simulation.

Table 2: Partition results of training and test samples.

Number	Number of risk identification training samples	Number of risk identification test samples
1	90	30
2	70	50
3	80	40
4	60	60
5	50	70

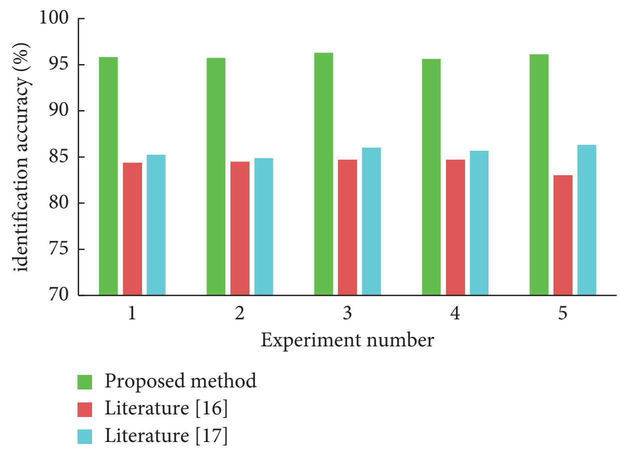


Figure 5: Financial information management system risk identification accuracy comparison.

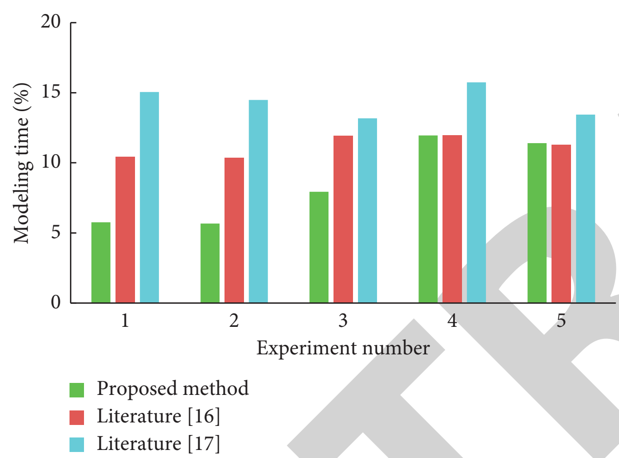


Figure 6: Comparison of risk identification modelling time of financial information management system.

overcomes the shortcomings of traditional methods and verifies the superiority of risk identification method of financial information management system based on data mining.

(4) The risk identification time of financial information management system based on data mining is less than that of [16] and [17], and the risk identification efficiency of financial information management system is significantly improved.

5. Conclusion

The research on enterprise financial management and prediction system is of great value. It has become a major subject of current financial research. The traditional method cannot describe the changing situation of financial information management system risk comprehensively and scientifically. They cannot guarantee the security of financial information management systems. In order to obtain the ideal risk identification effect of financial information management system, this paper proposes the research and design of enterprise financial management and prediction system based on SaaS model. Comparing with other risk identification methods of financial information management system, the results show that the financial management system proposed in this paper can objectively track the changes of financial information management system risk. The financial management system proposed in this paper establishes a risk identification model with higher accuracy. It shortens the risk identification time of financial information management system and can provide valuable reference information for financial information managers.

Data Availability

The labelled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This work is supported by the Hebei Oriental University.

References

[1] N. Tanklevska and V. Miroshnichenko, "Theoretical basis of managing of enterprises' financial potential on the basis of anti-crisis management," *Agricultural and Resource Economics: International Scientific E-Journal*, vol. 5, no. 2, pp. 51–61, 2019.

[2] T. Kliestik, K. Valaskova, K. Valaskova, G. Lazaroiu, M. Kovacova, and J. Vrbka, "Remaining financially healthy and competitive: the role of financial predictors," *Journal of Competitiveness*, vol. 12, no. 1, pp. 74–92, 2020.

[3] K. Goyal and S. Kumar, "Financial literacy: a systematic review and bibliometric analysis," *International Journal of Consumer Studies*, vol. 45, no. 1, pp. 80–105, 2021.

[4] H. Peng, "Research on credit risk identification of internet financial enterprises based on big data," *Mobile Information Systems*, vol. 2021, Article ID 1034803, 8 pages, 2021.

[5] A. Kim, Y. Yang, S. Lessmann, T. Ma, M.-C. Sung, and J. E. V. Johnson, "Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting," *European Journal of Operational Research*, vol. 283, no. 1, pp. 217–234, 2020.

[6] H. Zhou, G. Sun, S. Fu, J. Liu, X. Zhou, and J. Zhou, "A big data mining approach of PSO-based BP neural network for financial risk management with IoT," *IEEE Access*, vol. 7, pp. 154035–154043, 2019.

[7] S. Zhao, "Financial risk assessment and pre-warning of port enterprises based on neural network model," *Journal of Coastal Research*, vol. 110, no. SI, pp. 243–246, 2020.

[8] K. Valaskova, T. Kliestik, L. Svabova, and P. Adamko, "Financial risk measurement and prediction modelling for sustainable development of business entities using regression analysis," *Sustainability*, vol. 10, no. 7, p. 2144, 2018.

[9] M. Leo, S. Sharma, and K. Maddulety, "Machine learning in banking risk management: a literature review," *Risks*, vol. 7, no. 1, p. 29, 2019.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9892175, 1 page
https://doi.org/10.1155/2023/9892175

Retraction

Retracted: The Evaluation of DDoS Attack Effect Based on Neural Network

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Guo, H. Qiu, Z. Liu, J. Zhu, and Q. Wang, "The Evaluation of DDoS Attack Effect Based on Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 5166323, 16 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5166323, 16 pages
https://doi.org/10.1155/2022/5166323

Research Article

The Evaluation of DDoS Attack Effect Based on Neural Network

Wei Guo, Han Qiu, Zimian Liu, Junhu Zhu, and Qingxian Wang

State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450002, China

Correspondence should be addressed to Han Qiu; qiuhan_loach@sina.com

Received 2 March 2022; Accepted 22 March 2022; Published 11 April 2022

Academic Editor: Chin-Ling Chen

DDoS attack effect evaluation is the basis of security strategy deployment. The traditional effect evaluation method relies on the original data, ignoring the relationship between features and the evaluation target and indicator data redundancy, which affects the accuracy and reliability of the evaluation result. To this end, we introduce distance entropy to measure the similarity between features and evaluation target and use LSTM and Triplet networks to measure multiple correlations simultaneously. Then, a 2D-CNN is used to mine deep feature information and filter irrelevant information. We also combine 1D-CNN and attention models to achieve hierarchical sampling of different local features. Finally, three fully connected layers' training obtains a total evaluation value. We conducted experiments on five commonly used DDoS datasets. The results showed that the average ranking accuracy of the neural network-based DDoS attack evaluation method (NNDE) reached 87.2%, 91.3%, 88%, 85.6%, and 94.5%, respectively. Compared with other evaluation methods, an average increase of 19.73% indicates that this method can better evaluate the effect of DDoS attacks.

1. Introduction

The Distributed Denial of Service Attack (DDoS) has become one of the most threatening and lethal network attacks [1]. Its attack targets are usually critical Internet infrastructure or computing devices that provide essential services to outsiders. These devices include communication devices such as routers and gateways and high-performance servers that provide services such as education and banking. There are two characteristics among these intermediate nodes or terminals. One is to provide uninterrupted hardware support for diversified network operations and user access. The second is to verify the source of data packets by checking simple information such as source IP. Due to the lack of necessary traffic filtering methods, hackers can easily pretend to be ordinary users through IP spoofing [2]. They can also use reflection attacks to continuously send large-flow data packets to the target end to block normal services [3].

The low threshold and high yield of DDoS attacks make it possible for a fledgling novice to become an aggressive hacker overnight. Furthermore, their ignorance and aimlessness increase the frequency of cyberattacks. In order to cope with this network attack that may occur abruptly, the current intrusion detection system must constantly maintain a high alert state [4]. This comprehensive security strategy lacks threat grading. It treats all situations and periods as equally important, wasting a lot of extra labor and materials. Consequently, it is needed to evaluate the effect of DDoS attacks during an early attack period to distinguish the priority. The evaluation results can be used to purposefully assign scarce software and hardware resources to maximize defense gains.

The core of the effect evaluation is to design the corresponding evaluation algorithm according to the research question and the research object. An evaluation algorithm is a series of calculation steps that integrate multiple indicators into a total value. According to the input data, method attributes, and application category, we classified 15 classic evaluation methods in the DDoS evaluation field, as shown in Figure 1.

Univariate evaluation methods include index evaluation method (IEM) [5] and matrix analysis method (MAM) [6]. These methods are simple and have low computational

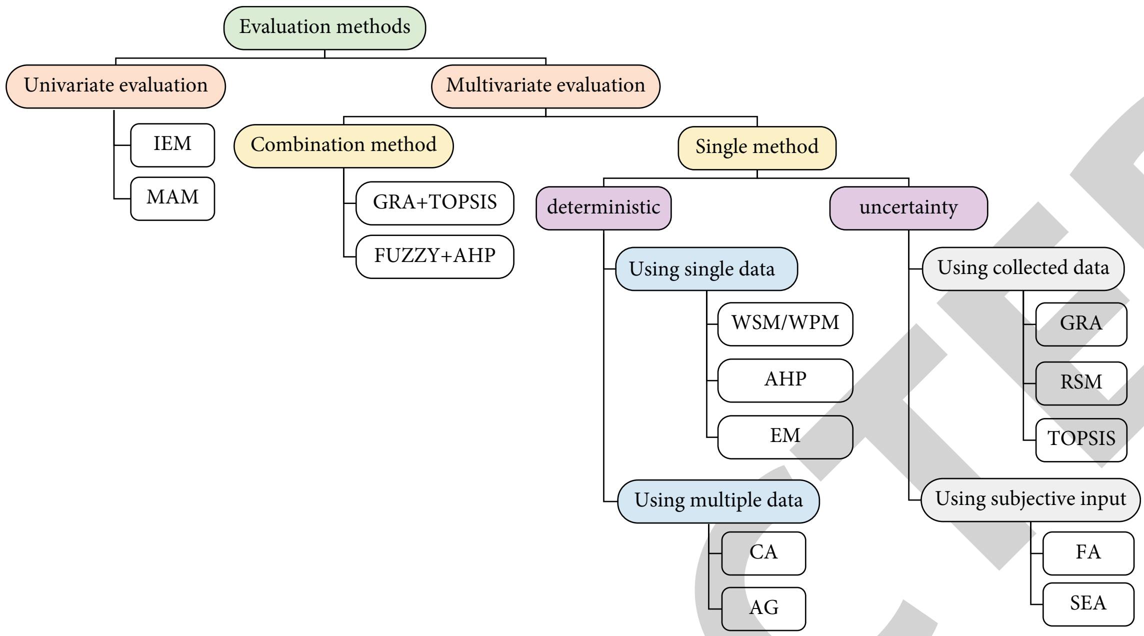


Figure 1: Classification of DDoS evaluation methods.

complexity. However, the one-sidedness of only one indicator determines that it cannot reflect all aspects of the evaluation object. Therefore, the multiattribute evaluation method is proposed to measure complex systems comprehensively. These multivariate evaluation methods include weighted sum/product method (WSM/WPM) [7], analytic hierarchy process (AHP) [8], and entropy method (EM) [9]. The above methods are all deterministic evaluation methods. They assume that all indicator data is available. When faced with the problem of uncertainty, the lack of ambiguous data conversion strategies makes these certainty methods unusable. Hence, uncertainty methods such as grey relational analysis (GRA) [10], rough set method (RSM) [11], and technique for order preference by similarity to an ideal solution (TOPSIS) [12] are proposed to quantify qualitative indicators. Besides, the evaluation method has also been improved to meet the needs of different applications. Deterministic evaluation methods such as WSM only require a set of data as input. It cannot make full use of the hidden relationships between different groups. In order to highlight the comparative significance of different results, evaluation methods that can process a numeric matrix in parallel have been proposed, including cluster analysis (CA) [13] and attack graph (AG) [14]. In uncertainty, manual data entry and configuration parameters are tricky issues. Aimed at this, fuzzy analysis (FA) [15] and system effectiveness analysis (SEA) [16] use fuzzy synthesis operators and trajectory domain to implement the fuzzy processing of system effectiveness.

The above evaluation methods provide calculation formulas that integrate DDoS evaluation indicators through mathematical theories such as statistics and topology. Nevertheless, there are still three significant problems. The first problem is that the result of the DDoS evaluation method is only determined by the input data and is not associated with the evaluation object. This one-sidedness leads to current methods, especially deterministic ones, susceptible to erroneous data and noise. Moreover, these interferences often appear in live network attacks, making this problem even more prominent. For example, the correlation coefficient of GRA and the ideal solution of TOPSIS are calculated based on extreme values, but whether each input is necessary has not been verified. The second problem is that the calculation of DDoS evaluation methods usually requires manual input of additional parameters, reducing the objectivity and credibility of the evaluation results. For example, the weight of WSM and the comparison matrix of AHP both need to be manually assigned. The third problem is that the calculation of the DDoS evaluation method defaults that all input indicators are independent of each other. These evaluation methods lack correlation processing, leading to redundant calculations in the evaluation results. For example, excessive memory usage will degrade software performance. The poor performance of the software makes the response delay of the provided service rise synchronously. Suppose that memory usage and response delay are directly integrated without correlation processing. In that case, the attack effect will be amplified, and the evaluation result will be distorted.

In the past ten years, artificial intelligence technology has made tremendous progress [17]. Significantly, the emergence of deep learning has dramatically accelerated its development, and various intelligent applications have emerged one after another [18]. This emerging technology has given us a new solution for evaluating the effectiveness of DDoS attacks. Deep learning is based on multilayer neural networks to train large sample sets. It can automatically learn the relationship between input and output. The neural network-based DDoS evaluation method has the nature of solving the above three problems. For question one, the evaluation no longer solely relies on the input data. The neural network connects input and output through a backpropagation algorithm. For question two, the neural network trains weights and bias values through gradient descent. The parameters are learned through training, thereby avoiding subjective weighting. For question three, the nonlinear activation function of the neural network can handle nonmonotonic relations. Moreover, structures such as the convolutional neural networks (CNN) implement feature conversion, so that the output variables remain independent of each other.

This paper proposes a new method for evaluating the effectiveness of DDoS attacks based on neural networks. First, the two relationships between features and targets are clarified by analyzing a typical DDoS feature generation tool. Further, feature selection is realized based on the combination of the Triplet network and long short-term memory (LSTM). Through the convolutional network, deep feature information can be mined. The improved attention model is used to grade different local features. Finally, the result is synthesized through the fully connected layer.

The method proposed in this paper is important because we are the first to study DDoS attack effect evaluation based on deep learning theory. The main contributions of this research are as follows:

(i) The deep learning architecture was used for the first time in the field of DDoS attack effect evaluation;

(ii) An improved triplet network based on distance entropy was used to achieve efficient feature selection for DDoS evaluation;

(iii) The introduction of a multihead attention mechanism solved the problem of assigning the same weight to different features.

The rest of this paper is organized as follows. Section 2 introduces background knowledge and related research. Section 3 extracts the DDoS characteristics and gives the neural network structure that can be used to evaluate the effect of DDoS attacks. Section 4 deals with the experimental dataset and verifies the effectiveness of our proposed neural network structure for DDoS evaluation. We sum up the research of this paper in Section 5. Section 6 gives the shortcomings of this paper and future research directions.

2. Related Work

In the DDoS effect evaluation, many evaluation methods are constantly being researched and put forward. Due to its low computational cost, the single indicator evaluation method is still widely used in DDoS evaluation. Paper [19] measured single indicators such as throughput and response time and analyzed the changes of indicators under different attack flow rates. Paper [20] compared the three entropy methods for evaluating the effect of DDoS attacks by analyzing the distribution of IP addresses. Paper [21] defined the amplification factor by calculating the proportion of delayed packets. It also implements corresponding reflection DDoS attack evaluation based on different application layer protocols. Paper [22] analyzed the impact on Tor through two typical DDoS attacks and measured the attack effect based on time consumption.

In the network confrontation environment, DDoS attack methods are usually diverse, and the background traffic is complicated. A single indicator evaluation method cannot meet the needs of comprehensive analysis in this case. Therefore, multivariant evaluation methods have gradually become the mainstream research direction of DDoS evaluation methods. Paper [23] proposed an automatic analysis of network attack effects based on attack graphs. This method assigns the probability of attack success to edge nodes to measure the degree of global damage. Paper [24] established a network attack indicator set and used fuzzy mathematics to give measurement criteria at different levels. Paper [25] established an evaluation indicator based on Delphi. It also used AHP to realize the security evaluation of the network system. Paper [26] established an indicator system by analyzing network attacks' benefit and cost functions. It also uses dynamic Bayesian networks to evaluate the impact of attacks on network nodes. Paper [27] proposed a new indicator weighting method and realized a multilevel feedback security evaluation system through the core algorithm.

To overcome the low credibility of a single evaluation method, scholars usually combine multiple methods to obtain more scientific evaluation results. Paper [28] combined the construction of the attack graph and the generation process of the Bayesian network to achieve different levels of network attack quantification methods. Paper [29] proposed a trust framework and defined the basic elements. In addition, the matrix analysis and clustering method were combined to achieve reliable DDoS evaluation. Paper [30] used the fuzzy AHP method to explain the relationship between the evaluation target and attributes and obtained the overall evaluation value based on TOPSIS. Paper [31] proposed a two-step evaluation method for DDoS attacks. First, use network traffic entropy to depict the distribution of data packets; second, merge the fuzzy idea with entropy to improve evaluation sensitivity.

By combining several evaluation methods, their advantages and strengths are inherited while avoiding defects, so that the accuracy and credibility of evaluation can be improved to a certain extent. However, how to choose the appropriate evaluation methods and how to achieve effective integration still require the subjective judgment of cybersecurity experts and data scientists. In addition, this strategy also needs to design various multiattribute fusion methods according to the changes of application scenarios. Its uncertainty limits the scope of application and cannot promptly respond to new types of attacks. The emergence of deep learning technology has given us new solutions for DDoS attack evaluation. This technique learns complex correlations between input and output variables through a multilayer neural network or perceptron and updates weights and biases through a backpropagation algorithm. In-depth correlation information is automatically extracted from the input features during this training process without human intervention. This automated procedure, which relies entirely on objective values, increases the reliability and confidence of the evaluation. In this paper, combined with the scene characteristics and elements of DDoS evaluation, we focus on the feature extraction and deep learning framework to design the DDoS evaluation method.

3. Methodology

To evaluate the effect of DDoS attacks, we extract features based on CICFlowMeter. In addition, we design a deep neural network to obtain the metric value of the DDoS attack effect. As shown in Figure 2, our proposed evaluation system consists of three main stages. First, the flow features in the collected original flow data are extracted as numerical vectors in the feature generation stage. On this basis, a feature subset suitable for DDoS attack evaluation is further selected. The training phase aims to establish an attack effect evaluation model. It takes the multidimensional feature vector at a certain moment as input. Then, output the evaluation value of the attack effect at this time. At this stage, a targeted neural network structure is selected to analyze the characteristics of DDoS traffic. Moreover, the number of the neural network's layers and the number of neurons in each layer are constrained. The accuracy measurement method of DDoS attack effect evaluation is proposed in the evaluation stage. Based on this, through comparison with other evaluation methods, the effectiveness of our proposed neural network-based DDoS attack effect evaluation method is verified.

3.1. Feature Extraction. This section first analyzes a commonly used DDoS feature generation tool to find ideas for optimizing data before training. Then, LSTM is used to capture the relationship of input features over time. Finally, we use the Triplet network to compute the two correlations between features and the evaluation object concurrently.

3.1.1. Feature Generation. CICFlowMeter is a commonly used tool for generating network traffic characteristics [32]. The well-known network security public datasets CIC-IDS2017, CIC-IDS2018, and CIC-DDoS2019 are collected based on this feature extraction tool [33]. Due to its extensive application and good evaluation performance, CICFlowMeter has become one of the most effective collection tools for DDoS feature generation.

This tool converts the collected pcap files into 77 measurable features according to the data distribution of the 5-tuple [source IP, source port, destination IP, destination port, transport protocol]. These 77 indicators can be divided into the flow, packet, mark, and segment according to the different fine-grained objects. The results are shown in Table 1.

The construction of the DDoS evaluation indicator system needs to meet two principles. One is that indicators should be strongly related to the evaluation object. For example, the flooding attack that uses large traffic as attack mode is a typical DDoS attack. This attack approach does not take advantage of the vulnerabilities in the three-way handshake protocol. The SYN or ACK sign in the acknowledgment packet is consistent with normal traffic. Therefore, the mark features "SYN Count" and "ACK Count" in Table 1 cannot reflect the attack effect of the flooding attack. If all indicators are used indiscriminately for effect evaluation, the results will contain more external interference. The second is to minimize the redundancy between indicators. From Table 1, we can see that similar features exist in a feature subset. For example, "Fwd IAT Std" and "Fwd IAT Mean" describe the flow interval, and "Packet Length Variance" and "Packet Length Mean" describe the packet length. There is a standard deviation and variance function relationship between them. Assume that this association is not considered. In this case, it will cause the implicit double calculation in the total value, affecting the accuracy of the final evaluation result.

The dataset labels can guide supervised training and measure the neural network's performance on the test set. DDoS attack effect evaluation belongs to supervised learning, so a good label is crucial. CICFlowMeter only emphasizes basic flow characteristics and ignores the generation of corresponding labels. Paper [34] defines a new label generation method. It is proposed that, under a sufficiently large sample, the attack effect can be estimated by the attack duration and the number of data packets. DDoS attacks have enormous data packets, which satisfies the premise. More importantly, in addition to the forward attack flow directly sent by the attacker, the DDoS attack also includes the backward attack flow generated by the reflective amplifier. Therefore, we utilize the three features "Flow duration," "Total Fwd Packet," and "Total Bwd Packet" to calculate the attack effect label. Its calculation formula is as follows:

$$\text{label} = \text{duration} \cdot \text{packet}_{\text{fwd}} \cdot \text{packet}_{\text{bwd}}. \quad (1)$$

We do not use equation (1) but a comprehensive algorithm to measure the effect of DDoS attacks. The reason is that, in a statistical sense, equation (1) can only be approximated when there are enough sample points. It would lose efficacy in a small interval. In the early attack stage, realizing quick attack effect measurement based on small data can gain more prime time for subsequent security strategy deployment. Therefore, we aim to find a way to quickly measure small samples' attack effect through a comprehensive analysis of large samples (such as neural networks).

3.1.2. Feature Selection. Feature selection is a necessary prestep for efficient neural network training [35]. According to Section 3.1.1, CICFlowMeter can generate a complete set of DDoS features. Dataset optimization can be achieved by connecting feature selection afterward. It achieves the filtering of irrelevant features by adjusting the objective function. It also reduces the probability of redundancy by controlling the number of features and speeds up the convergence of the neural network. The current DDoS feature selection depends on the empirical judgment of

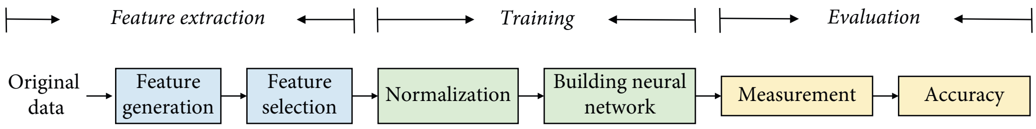


Figure 2: The overall process of DDoS attack effect evaluation based on neural network.

Table 1: Classification based on 77 features collected by CICFlowMeter.

Sort	Features
Flow	Flow duration, Flow Bytes/s, Flow Packets/s, Flow IAT Mean/Std/Max/Min, Fwd/Bwd IAT Min/Max/Mean/Std/Total, Down/Up Ratio.
Packet	Total Fwd/Bwd Packet, Total Length of Fwd/Bwd Packet, Fwd/Bwd Packet Length Min/max/Mean/Std, Fwd/Bwd Packets/s, Packet Length Min/Max/Mean/Std/Variance, Averge Packet Size.
Mark	Fwd/Bwd PSH/URG Flags, Fin/SYN/RST/PSH/ACK/URG/CWR/ECE Flag Count.
Segment	Fwd/Bwd Header Length, Fwd/Bwd Segment Size Avg, Fwd/Bwd Bytes/Bulk Avg, Fwd/Bwd Bulk Rate Avg, Subflow Fwd/Bwd Packets/Bytes, Fwd/Bwd Init Win bytes, Fwd Act Data Pkts, Fwd Seg Size Min, Active/Idle Min/Mean/Max/Std.

network security administrators. This method is not objective, easy to prioritize nonserious cyber attacks due to inevitable subjective errors. Consequently, a vicious attack will lose the best time to handle it.

Let *st* denote the relationship function between features and objects, and let *re* denote redundancy between features. Then, the best indicator system can be expressed as follows:

$$\text{Best\_subset} = \arg_{\text{subset}}(\max(st(\text{subset})), \min(re(\text{subset}))). \tag{2}$$

Triplet network (TN) is an improved version of the Siamese network (SN) [36, 37]. SN implements the similarity measurement of two features in the same set of inputs. It cannot calculate the relationship among three or more features. TN expands SN's two parallel sample inputs to three, thereby modeling the relationship between multiple features. The three inputs can be expressed as {anchor sample x, positive sample x^{+}, and negative sample x^{-}}. Unlike conventional neural networks that rely on labels, this model is an unsupervised learning network. Its training depends on the comparison between features. A set of feature inputs is essentially a time vector, and its time variation pattern can be obtained through LSTM [38]. Inspired by the TN model, we proposed an improved model, namely, LSTM based Triplet network (LTN), to achieve multiple correlation analyzes. We use positive sample x^{+} to simulate object *label*, anchor sample x to simulate candidate indicator i, and negative sample x^{-} to simulate remaining indicator $\bar{i}$. The LTN model is shown in Figure 3.

In Figure 3, dataset S is a real matrix with m features as rows and n moments as columns. Among inputs of LTN, feature set $i = \{i_1, i_2, \ldots, i_m\}$, time set $t = \{t_1, t_2, \ldots, t_n\}$, and d_{pq} denotes the value of feature i_p collected at time t_q. label = $\{l_1, l_2, \ldots, l_n\}$ is the ideal value for evaluating the effect of DDoS attacks at different times. Specifically, feature set i can be collected by a feature generator tool, such as CICFlowMeter, or designed subjectively by researchers. Time set t is the collection time interval sequence. Object label can be obtained by mapping existing labels in the public dataset or approximated by equation (1). Among the three neural network inputs in Figure 3, evaluation object label is the reference effect values, fixed as the label sequence unchanged. Candidate indicator i is randomly assigned from the feature set. Remaining indicator $\bar{i}$ is selected from other features except i. Notably, the disorder indicator pair $\langle i, \bar{i}\rangle$ cannot be calculated repeatedly. Thus, there are a total of C_m^2 indicator pairs. After C_m^2 iterations, the feature selection is completed.

The three features simultaneously input the same neural network structure and share weights and bias values in one iterative calculation. Each feature is trained separately through a neural network and does not interfere with each other. The input indicator is essentially a long-time series vector. Therefore, to find the complex relationship between different moments, we use a three-layer LSTM as the neural network structure. In order to coordinate multiple iterations, a triplet loss node is added at the end of LTN. This node updates the gradient descent process through backpropagation. Triplet loss aims to achieve the smallest redundancy between features and the strongest association with the target concurrently in multiple iterations. For triplet $(\text{label}, i, \bar{i})$, the calculation formula of triplet loss is expressed as

$$\text{Triplet\_Loss} = \max(d(i, \text{label}) - d(i, \bar{i}) + \ell, 0), \tag{3}$$

where $d(i, \text{label})$ and $d(i, \bar{i})$ are the Euclidean distances from candidate indicator i to object *label* and another indicator $\bar{i}$, and ℓ is the minimum distance between these two distances. Empirically speaking, ℓ is usually set to 0.6 [39], and $d(*, *)$ uses the softmax function to convert its value between 0 and 1. However, this distance is based on the solution of the spatial distance. It represents the linear distance between two sample points in a multidimensional space. It cannot reflect the coordinated change of the two. Aiming at this problem, we introduce the distance entropy of order α to mirror the difference in the overall system distribution. This distance can measure monotonic correlation. The calculation formula of $D(i, \text{label})$ is as follows:

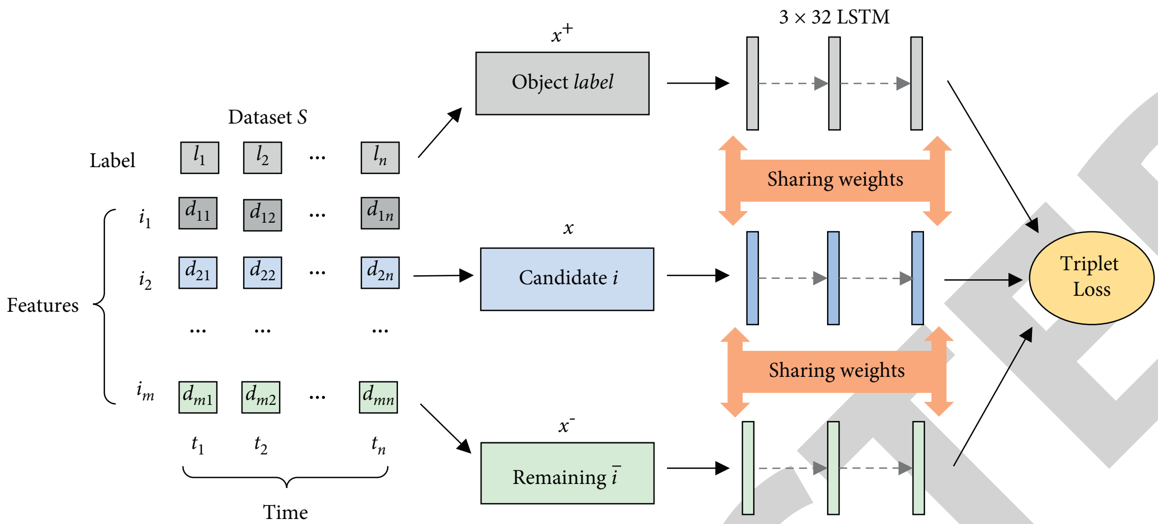


Figure 3: DDoS attack feature selection model LTN.

$$D(i,\text{label}) = \frac{1}{\alpha-1}\log_2\left(\sum_n\left(\frac{d_{mn}}{\sum_n d_{mn}}\right)^{\alpha}\left(\frac{\text{label}_n}{\sum_n \text{label}_n}\right)^{1-\alpha}\right). \tag{4}$$

Similarly, the calculation formula of $D(i,\bar{i})$ is as follows:

$$D(i,\bar{i}) = \frac{1}{\alpha-1}\log_2\left(\sum_n\left(\frac{d_{mn}}{\sum_n d_{mn}}\right)^{\alpha}\left(\frac{\bar{i}_n}{\sum_n \bar{i}_n}\right)^{1-\alpha}\right). \tag{5}$$

In equations (4) and (5), α is set to 6 empirically [20]. Combining these two, the formula of triplet loss entropy is as follows:

$$\text{Triplet\_Entropy} = \begin{cases} D(i,\text{label}) - D(i,\bar{i}) + \ell, & D(i,\text{label}) - D(i,\bar{i}) > \ell, \\ 0, & D(i,\text{label}) - D(i,\bar{i}) \le \ell. \end{cases} \tag{6}$$

The classification result based on equation (6) is used to judge whether a feature is suitable for DDoS effect evaluation.

3.2. Structure of Deep Learning Model. We proposed a novel neural network-based DDoS evaluation (NNDE) to measure the DDoS attack effect. It includes four main parts, namely, embedding layer, convolutional layer, attention model, and fully connected layer. The overall structure of NNDE is plotted in Figure 4. Besides, we explain each part's selection basis and implementation details in the following subsections.

3.2.1. Embedding. The initial feature set is obtained by simplifying the original dataset according to the feature selection model in Section 3.1.2. Compared with the original dataset, the number of features in the initial dataset is reduced. Each feature is a numeric vector collected under a multidimensional time series. What is more, the long-term confrontation of DDoS attacks makes this time series a high dimension. For example, in the CIC-DDoS2019 test set, the traffic items of TFTP reached 20,107,827. High-dimensional data makes DDoS evaluation a curse of dimensionality [40]. It will be challenging to converge if it is directly input to the feedforward neural network without further data processing. Therefore, to facilitate subsequent neural network training, the reduced feature set will first pass through an embedding layer to reduce dimensionality. Each state vector of the embedding layer is obtained by calculating the relationship between two features in the feature set. This relational calculation can be mapped from a high-dimensional sparse matrix to a low-dimensional dense matrix. All vectors in the embedding layer constitute the embedding matrix. When the matrix size is 64×1000, it can meet the dimensionality reduction requirements of most network security datasets.

3.2.2. Convolutional Layer. The convolution part comprises a two-dimensional convolutional neural network (2D-CNN) and a max-pooling layer. The output of the embedding layer is still a numerical matrix, so we use 2D-CNN rather than a one-dimensional convolutional neural network (1D-CNN) for training. The 2D-CNN layer has two functions. First, the

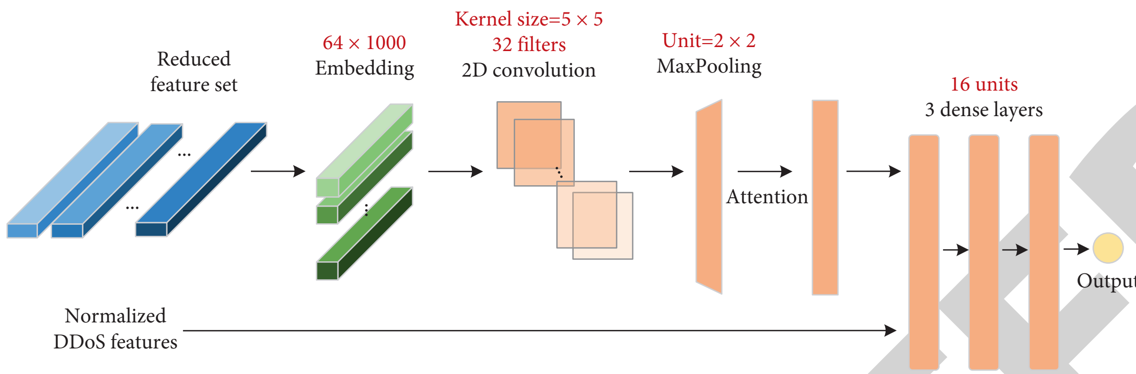


FIGURE 4: The deep learning structure of NNDE for DDoS effect evaluation.

translation of the convolution kernel is used to highlight local features. During the shifting process, insignificant information is ignored, thus speeding up the parallel calculation in the neural network. Secondly, the mutual relationship between input features and the implicit relationship between features and targets are extracted through the cross-correlation operation. In this process, initial standardized features are generated. In a denial of service attack, usually, no more than ten characteristics reflect the performance of a particular attack. Considering translation efficiency and fine-grained operations, we set the convolution kernel's size to 5×5, and its number to 32.

The same attack methods and patterns may be repeated in long-term DDoS attacks. This model leads to a similar distribution of sample points in different periods. Convolution operation has strict edge output, difficult to merge similar data. Therefore, we added a max-pooling layer after the 2D-CNN layer to reduce the sensitivity to the sampling position.

3.2.3. Attention Model. The convolutional layer focuses on the weight of a single feature. However, it ignores the comparison of local features, so that the learned different local features have the same importance. Consequently, the impact of multiple local features on the network attack effect cannot be well measured. We introduced a multihead attention model to add weight to different local features and eliminate the influence of repeated features [41]. The structure of the multihead attention model is shown in Figure 5. The idea of attention comes from the automatic coder-decoder [42], which calculates the input of the next moment through the weighted average of the implicit correlation between current outputs. We used the multihead attention mechanism to capture the interaction between different local features. We also modified the attention model to better deal with DDoS characteristics. We added ResNet to reconstruct the intermediate data and 1D-CNN for information clustering and parameter reduction [43]. The reason for choosing 1D-CNN here is that the feature matrix is segmented and separately calculated after ResNet and Add&Norm.

In Figure 5, the attention model includes three main parts: multihead attention, residual and normalization, and 1D-CNN. Among them, multihead attention integrates the deeply hidden information learned by the model in multiple expression subspaces [44]. Let i^u be the u-th indicator in the indicator set i. Take the parameter transfer process of i^u in each subspace to illustrate information flowing in the attention model. Suppose that the total number of indicator heads is R, then arbitrary indicator head $r \in \{1, 2, \ldots, R\}$. Let $\delta^r(i^u, i^v)$ denote the correlation between indicator u and indicator v, and let $\eta^r_{u,v}$ denote the correlation coefficient between indicator u and indicator v. The relationship between the two is as follows:

$$\eta^r_{u,v} = \frac{\exp\left(\delta^r\left(i^u, i^v\right)\right)}{\sum_{j=1}^{m} \exp\left(\delta^r\left(i^u, i^j\right)\right)}. \tag{7}$$

In equation (7), the calculation method of $\delta^r(i^u, i^v)$ includes dot product and perceptron. The dot product has a less computational cost, so we use it to converge quickly. Its formula is as follows:

$$\delta^r\left(i^u, i^v\right) = \langle Q^r i^u, K^r i^v \rangle, Q^r, K^r \in \mathbb{R}^{s', s}, \tag{8}$$

where Q^r and K^r represent the key-value pair $\langle \text{Query}, \text{Key} \rangle$ of the input indicator i^u. This key-value pair is the conversion parameter that maps indicator i^u from the initial sample space $\mathbb{R}^s$ to the new sample space $\mathbb{R}^{s'}$. The essence of equation (8) is to calculate the correlation between the queue and key. Its result is used as the attention weight. The calculation formula for the indicator head r of the indicator i^u is as follows:

$$\tilde{i}^u_r = \sum_{j=1}^{m} \eta^r_{u,j}\left(V^r i^j\right), V^r \in \mathbb{R}^{s', s}. \tag{9}$$

The mapping of the indicator i^u in the new sample space can be formed by concatenating all the conversion formulas $\tilde{i}^u_r$ in the indicator header set. The result is as follows:

$$\tilde{i}^u = \text{Concat}\left(\tilde{i}^u_1, \ldots, \tilde{i}^u_R\right). \tag{10}$$

Up to now, we have regenerated the weights of different local features. In the above process of generating the new space solution, the calculation of different indicator heads is parallelized. However, it is prone to decay in training due to more network layers and is difficult to converge. To solve this, we use ResNet to optimize the network structure. It makes the training process more sensitive to parameter

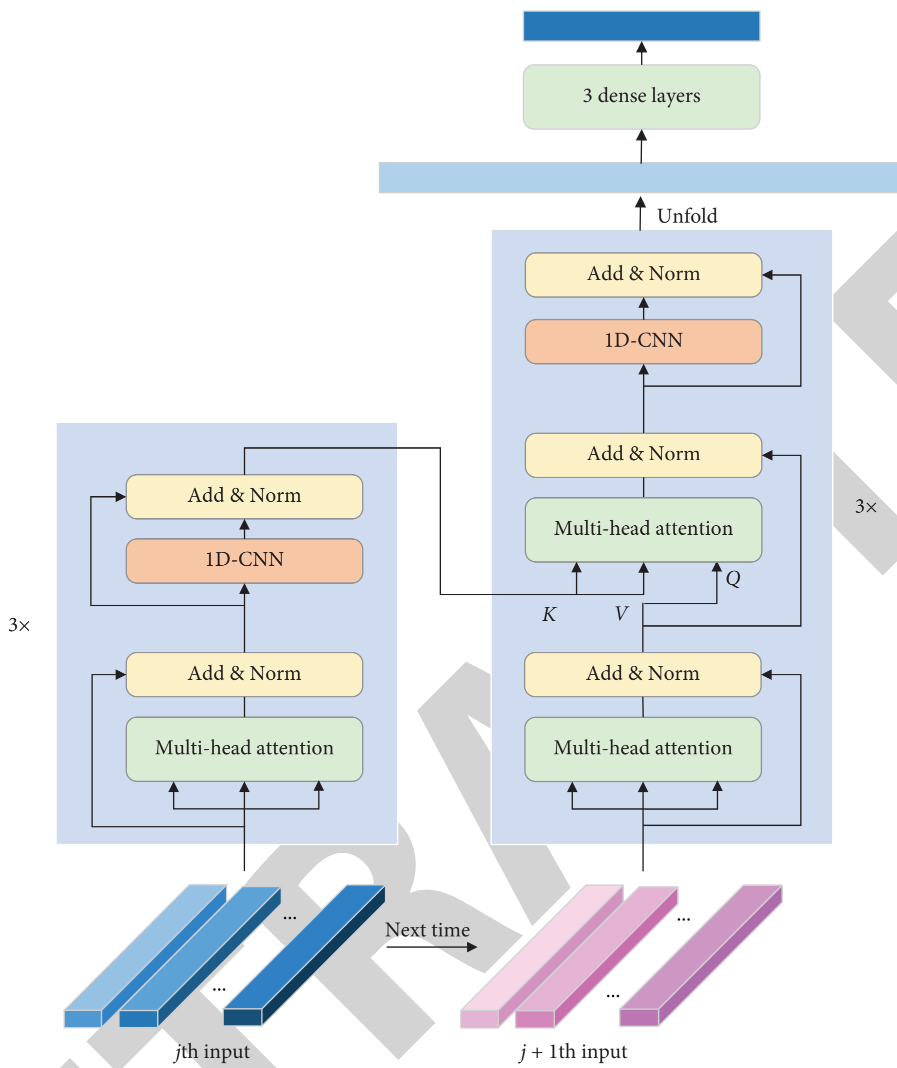


Figure 5: The structure of the attention model and fully connected layer.

changes, improves the efficiency of forward propagation, and solves the disappearance of backward propagation gradients. After N-layer ResNet, the change of indicator i^u can be expressed as

$$\tilde{i}^u = \tilde{i}^u + \sum_{j=1}^{N-1} \mathrm{ReLu}\left(\tilde{i}^j, w^j\right), \tag{11}$$

where $w^j \in \mathbb{R}^{s,s}$ denotes a transform matrix composed of the jth layer network's parameters, and ReLu is a nonlinear activation function. After ResNet, they are connected to the normalization layer to make the gradient descent training process smoother and improve the generalization ability. Under the output $\left\{\tilde{i}^j, 1 \le j \le m\right\}$ of different features, the overall output $\tilde{T}$ of the attention model can be expressed as

$$\tilde{T} = \mathrm{Concat}\left(\tilde{i}^1, \ldots \tilde{i}^m\right). \tag{12}$$

After pooling and ResNet, the intermediate data structure has changed. Thus, we utilize 1D-CNN to remine the vague association relationship between the input features. Compared with the feedforward neural network, 1D-CNN has fewer middle parameters, thus speeding the calculation. Subsequent experiments have shown that it is better when the attention base module is executed three times. Under this parameter, assigning different weights to different features will achieve the best balance between operational efficiency and evaluation accuracy.

3.2.4. Fully Connected Layer. The fully connected layer uses neurons to perform nonlinear transformations on the output of the attention model to obtain the predicted value of the DDoS attack effect. Commonly used nonlinear activation functions include sigmoid, tanh, and ReLu [45]. Among them, the asymmetric expression of sigmoid makes its output not centered on the origin. This deviation causes the weight to tilt to the positive axis during the deep gradient descent, affecting the comparative meaning of the difference in attack effect evaluation at different points. Tanh takes exponential calculation as the core. Its backpropagation slope is low and tends to the problem of gradient disappearance, which leads to a slow learning process. ReLu takes the origin as the center of rotation and alleviates the interference of non-zero symmetry on the result to some extent. This function is based on the linear operation, so the calculation rate is fast. Further, the oversaturation problem can be effectively solved by adjusting the learning rate. Therefore, we use ReLu as the activation function.

The number of fully connected layers usually appears as a hyperparameter. Besides adjusting the parameters during the experiment, we also try to explain the meaning of the number of layers theoretically. For this reason, we compared the application scenarios, advantages, and disadvantages of fully connected networks with different layers, as shown in Table 2.

DDoS attack effect evaluation is a regression problem. From Table 2, we can see that the regression solution requires that the number of fully connected layers is not less than two. In addition, its number is not less than three to learn deeper complex features. However, converging in more than four fully connected layers is challenging. Combining the above restrictions, we set the number of fully connected layers to three, each with 16 units. Unlike the classification problem, the activation function of the last layer is not softmax but sigmoid, which generates an actual value between zero and one. Its size indicates the effectiveness of the attack. A dropout layer is connected behind the convolutional and fully connected layers to prevent overfitting. Let w and b denote the weight and bias of the fully connected layer, respectively. The final attack effect value at time t can be expressed as

$$T_t = \mathrm{Sigmoid}(W\widetilde{T} + b), W \in \mathbb{R}^{1\times m\times s}. \quad (13)$$

4. Evaluation

In the field of network security, commonly used DDoS datasets include KDD99 [46, 47], NSL-KDD2009 [48], CIC-IDS2017 [49], CIC-IDS2018 [50], and CIC-DDoS2019 [51]. The KDD99 and NSL-KDD2009 datasets are usually used as benchmark datasets for DDoS attack effect evaluation. The CIC-IDS2017, CIC-IDS2018, and CIC-DDoS2019 datasets have overcome the shortcomings of the benchmark datasets, such as unextracted reverse features and uncaught reflection DDoS attacks. These three datasets have become the most convincing representatives of network security datasets. To improve the experiment's credibility, we chose these five well-known DDoS public datasets as the experimental datasets.

The experimental part includes the validity verification of feature selection and the validity verification of the proposed neural network. The design logic is shown in Figure 6. It can be seen from Figure 6 that the necessity and effectiveness of feature selection are first verified. Secondly, the effectiveness of DDoS attack evaluation based on the proposed neural network is verified under the premise of feature selection. The comparative experiment includes four basic units: dataset, neural network, evaluation algorithm, and feature selection method. The connection between these four units and an experiment indicates that they will be considered in the current experiment. For example, the link marked ① connects the neural network and evaluation dataset to the first experiment, indicating that this experiment needs to consider the influence of different datasets and network structures.

Specifically, in the comparison experiment of feature selection methods, we selected three standard statistical feature selection methods and four classical neural networks as the comparison objects. Three statistical methods include chi-square (CHI) [52], relevant feature selection (Relief) [53], and information gain (IG) [54]. The four classic neural network structures include deep neural networks (DNN), CNN, recurrent neural networks (RNN), and CNN + RNN. Besides the neural network structure, four common evaluation methods are selected as the comparison objects in the comparison experiment of the attack effect evaluation, namely, certainty evaluation WSM, uncertainty evaluation GRA and TOPSIS, and combined evaluation GRA + TOPSIS.

It is worth mentioning that, for clarification, we did not compare the method proposed in this paper with popular DDoS detection methods. As we all know, both DDoS detection and DDoS assessment can be analyzed and studied through deep learning methods. However, in essence, DDoS detection is a classification problem, and DDoS evaluation is a prediction problem. There are great differences between the two in terms of research purpose, realization method, and parameter design. For example, DDoS detection distinguishes attack traffic from normal traffic through traffic characteristics, and DDoS evaluation quantifies the strength of an attack by integrating evaluation indicators. The former finally gets 0 or 1 to mark whether it has suffered a DDoS attack, while the latter's calculation result is a scalar that can take an arbitrary floating-point number, which represents the strength of the attack the defender is experiencing. This scalar contains additional information that cannot be expressed simply by solving the binary classification problem. Therefore, to highlight the difference between the two, we did not compare the state-of-the-art methods for DDoS detection with the NNDE evaluation method proposed in this paper.

4.1. Data Processing. First, we extracted DDoS-related attack traffic from five datasets. The number is shown in Table 3. Notably, the NSL-KDD2009 dataset has added the defense difficulty label. The larger this number, the easier it is to detect the attack, and the lower the attacking threat, and conversely, the higher the threat of attack. This label can be used to indicate the attack effect. Differently, the other four datasets are unlabeled. So, we manually added the label according to Eq. 1.

Noteworthily, neural networks and evaluation algorithms require numeric vectors as input and cannot handle string data. These string data include two types, namely, numbers with no comparative meaning and undefined special characters. For the first type, such as traffic ID and source IP, we deleted them from the dataset.

Table 2: Comparison of different fully connected layers.

Num	Applicable scene	Disadvantages	Advantages
0	Linear, simple classification	Cannot cope with complex relationships	Simple, fast convergence
1	Nonlinear, simple classification	Low accuracy	Continuous function can be expressed
2	Nonlinear, classification and regression	Limited learning characteristics	Smooth mapping can be expressed
3	Nonlinear, classification and regression	Slow convergence	Complex features can be learned
≥4	Nonlinear, classification and regression	Very slow convergence	Complex features can be learned

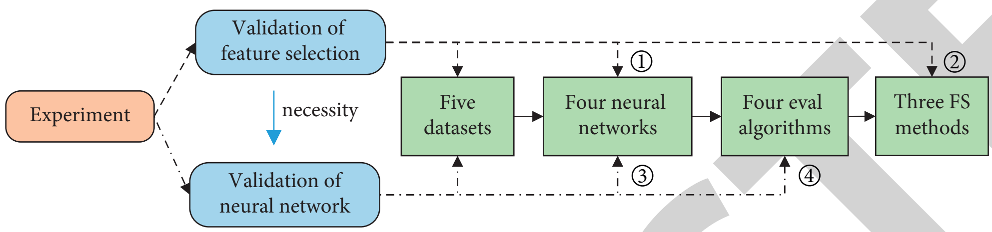


Figure 6: Two of the validation subexperiment design logic.

Table 3: The number of DDoS-related traffic in five datasets.

Dataset	KDD99	NSL-KDD2009	CIC-IDS2017	CIC-IDS2018	CIC-DDoS2019
DDoS traffic	4856151	45927	128027	687742	50063112

For the second type, we replaced "Infinity" with statistical values (such as extremum) and removed traffic data "NaN."

Since the dimensions and range of indicators are different, it is senseless to synthesize them directly. So, indicators must be standardized before the evaluation begins. There are three main normalization methods: StandardScaler, MinMaxScaler, and RobustScaler [55]. Among them, the calculation of mean and standard deviation in StandardScaler and the extreme value in MinMaxScaler are easily affected by outliers, affecting the comparison of intermediate values. In contrast, RobustScaler uses quantiles as a reference, so a single value does not affect this method. To sum up, we chose RobustScaler as the normalization method to reduce errors. Its formula is as follows:

$$d'_{pq} = \frac{d_{pq} - \text{median}}{\text{quantile}_{75} - \text{quantile}_{25}}, \tag{14}$$

where median denotes the mean value, quantile_{75} denotes the upper quartile, and quantile_{25} denotes the lower quartile.

4.2. Measurement. The validity verification of our proposed evaluation method includes the comparison with mainstream neural networks and commonly used evaluation algorithms. In the comparative experiment of neural networks, the minimum mean square error (MSE) [56] does not signify that the evaluation value is the most accurate. It is also essential to assess whether the results are evenly distributed. Therefore, in addition to MSE, we also use the explained variance score (EVS) [57] to investigate the discrete distribution of results. Their formulas are as follows:

$$\begin{aligned} \text{MSE} &= \frac{1}{n}\sum_{j=1}^{n}\left(T_j - \text{label}_j\right)^2, \\ \text{EVS} &= 1 - \frac{\text{Var}\left(T_j - \text{label}_j\right)}{\text{Var}\left(T_j\right)}, \end{aligned} \tag{15}$$

where n denotes the sampling time, T_j and label_j denote the estimated value and label value at the jth time, respectively, and $\text{Var}(*)$ denotes the variance calculation. Because the calculation principles are not the same, different evaluation algorithms will usually get distinct values for the same data input. In the comparison experiment with other evaluation algorithms, common measures MSE and MAE were not selected. The reason is that both calculate the spatial distance between single points and lack comparison. Further, we put forward a new measure Sort_Acc to compare effects at different moments. It is defined as follows:

$$\text{Sort\_Acc} = \frac{\text{num}\left(i^u \cap \text{label}\right)}{C_n^2}, \tag{16}$$

where $\text{num}(i^u \cap \text{label})$ denotes the intersection of ordered pairs in i^u and label, and C_n^2 denotes the number of all possible ranking pairs. In equation (16), there are three relationships: greater than, less than, and equal to. The nonequal relationship between two labels is judged whether the comparison relationship of the corresponding predictive values is consistent with it. If equal, Sort_Acc will increase by one. The equality relationship between two labels is judged whether the difference between the corresponding outputs exceeds 10% of the smaller value. If less, the two values are equal, and Sort_Acc will increase by one. Otherwise, it is not.

In this part, according to the different comparison objects, we choose different measurement strategies to illustrate the quality of the attack effect evaluation. When the proposed NNDE method is compared with other neural network methods, the commonly used metrics MSE and EVS can cope with this situation. However, the same input usually results in greatly different output values for different statistical evaluation methods due to the variable calculation principles. In this case, MSE or EVS based on numerical regression cannot be used for comparison. Although the obtained values are different, the order of the different evaluation values can still reflect the quality of the evaluation results. Based on this, when comparing with statistical methods, we proposed a new metric named Sort_Acc to uniformly describe the impact of evaluation methods on ranking evaluation results.

4.3. Result and Analysis. We built a Keras-based neural network test platform to verify the effectiveness of our proposed method. TensorFlow is used as the backend training engine. It supports modular network structure programming and can execute on a CPU or GPU. What is more, the parallel computing capabilities of GPUs have greatly improved the training speed of deep neural networks.

The configuration parameters of the workstation running the experiment are 64G RAM, 1TB solid-state disk, i5-9300HF processor, and GeForce GTX 1660Ti graphic card. Since the number of normal samples in the dataset is much larger than the number of attack samples, we use k-fold cross-validation to divide the original dataset. Concretely, in the experiment, the value of k is set to 10. In each training, one share is selected as the test set, and the remaining nine shares are used as the training set.

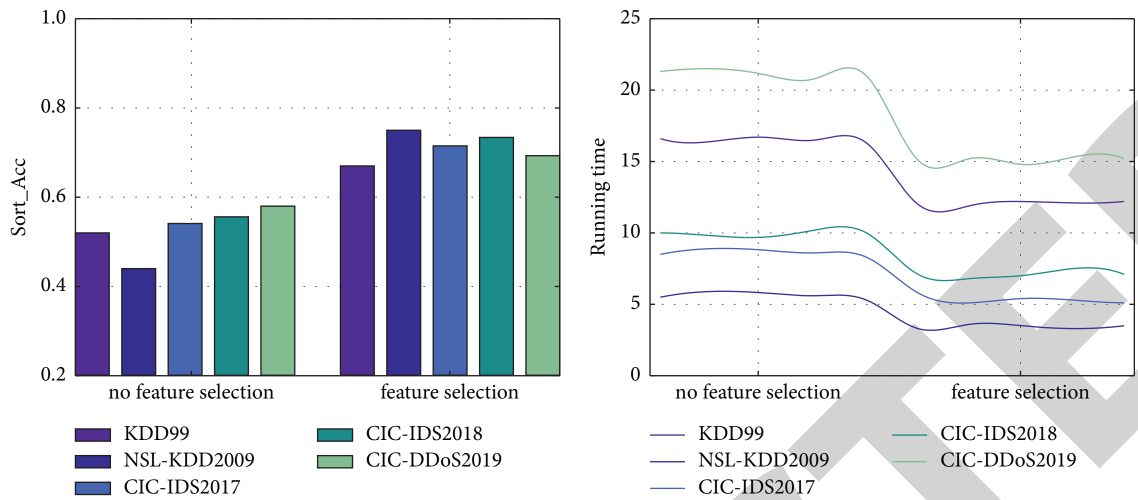
4.3.1. Validation Experiment of Feature Selection. First, we examined the necessity of feature selection for DDoS attack effect evaluation. No feature selection was used as the control group, and feature selection was used as the experimental group. We compared the difference between the two in terms of accuracy and running time. Specifically, CHI, Relief, and IG are used as the three feature selection methods of the experimental group, and CNN and RNN are used as the two evaluation neural networks. Train ten times and take the average to reduce errors. The verification results on the five datasets are shown in Figure 7. It can be seen from Figure 7(a) that the DDoS attack effect evaluation with feature selection has achieved a higher accuracy rate than that without feature selection. Especially on the NSL-KDD2009 dataset, the accuracy rate after feature selection increases from 44.1% to 75.3%, and the improvement rate reaches 31.2%. Figure 7(b) shows that feature selection can remarkably reduce the training time. When the amount of data is larger, the decrease rate in time consumption is greater. On the CIC-DDoS2019 dataset, the highest average consumption time reduction rate is 28.5%. In short, feature selection can improve training efficiency and training accuracy at the same time and is an essential part of efficient and high-performance effect evaluation.

Secondly, the effectiveness of our proposed LTN model was verified by comparing it with other neural network structures and statistical feature selection methods. We reproduced the traditional feature selection methods and ran them on the CPU. Deep learning methods were run on GPU and CPU separately to compare time consumption. The comparison includes the accuracy of ranking pairs and the training time of evaluation. The average of multiple pieces of training is taken. K99, N09, C17, C18, and C19 represent five datasets. The results are shown in Table 4. It can be seen from Table 4 that the LTN method has reached the highest sorting accuracy on the five datasets, respectively, 93.1%, 92.3%, 92.9%, 93.7%, and 94.6%. The statistical-based feature selection method has the lowest accuracy rate, and its highest accuracy rate does not exceed 74%. In addition, the combination of neural networks has achieved better performance than a single network. However, there is still an average gap of 5.52% from the LTN method. From Table 4, we can see that, in addition to DNN, deep learning methods converge faster than traditional statistical methods. In particular, the LTN method achieves the fastest convergence rate on five datasets. The deep learning method running on GPU during the training process is about 16 times more efficient than the CPU. The DNN structure has many parameters, and statistical methods cannot use GPU acceleration, making it difficult for both to process large-scale data.

All in all, the comparison with the other seven feature selection methods verifies the effectiveness of the LTN method in selecting DDoS evaluation features. This method can better reduce the training time and improve evaluation accuracy. In subsequent experiments, the feature selection method is set to LTN.

4.3.2. Validation Experiment of Neural Network Evaluation. First, we verified the effectiveness of NNDE for DDoS effect evaluation under the comparison of four neural networks. Specifically, the single DNN or CNN structure maintains the same parameter settings as the NNDE network to control variables. For RNN, we use a three-dimensional LSTM structure of 32 neurons to represent it. MSE and EVS are used to measure neural network regression error and distribution, respectively. The comparison result is shown in Figure 8. It can be seen from Figure 8 that the NNDE method achieves the smallest regression error and the highest distribution stability on the five datasets. From Figure 8(e), our proposed NNDE method performs best on the CIC-DDoS2019 dataset. Its average error is 0.27, and the variance score is 0.82. In particular, as shown from Figure 8(b), on the NSL-KDD2009 dataset with small samples, NNDE still controls the average error within 0.45, and the variance number exceeds 0.7, showing a robust fitting ability.

Secondly, we also verified the effectiveness of NNDE under the comparison of four classic evaluation algorithms. The python-based Pandas library implements the parallel matrix calculation of the evaluation algorithm. Specifically, WSM averages multiple weight calculations

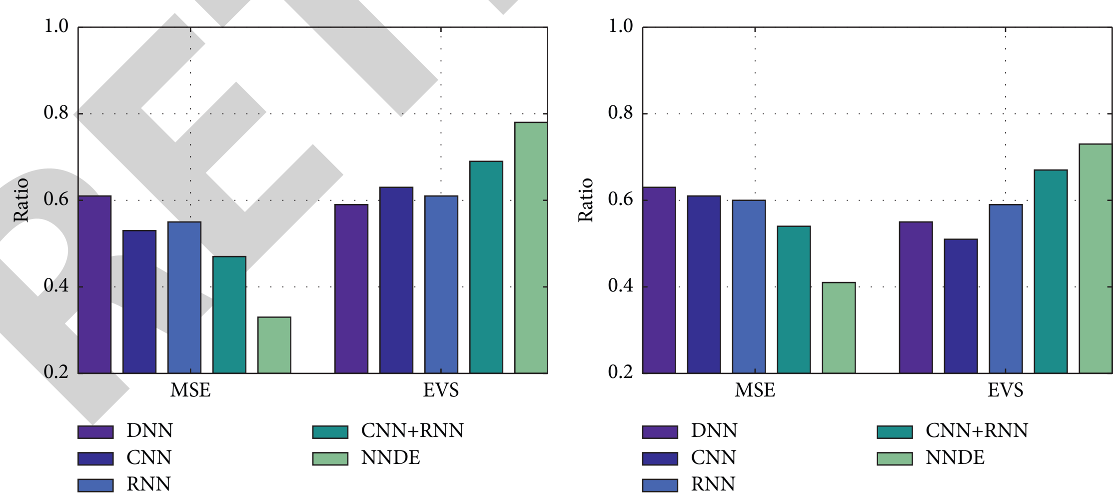


(a) (b)

Figure 7: The necessity of DDoS feature selection in effect evaluation.

Table 4: Comparison of LTN and other feature selection methods.

FS method	Sort_Acc					Running time				
	K99	N09	C17	C18	C19	K99	N09	C17	C18	C19
DNN	0.768	0.797	0.826	0.802	0.811	56.41	25.11	31.61	40.32	2482.29
CNN	0.813	0.785	0.812	0.834	0.835	19.62	14.43	15.35	16.88	789.53
RNN	0.79	0.763	0.829	0.836	0.842	21.14	16.06	16.99	20.71	861.81
CNN + RNN	0.877	0.854	0.871	0.874	0.914	17.06	14.26	14.86	19.98	738.9
CHI	0.671	0.652	0.664	0.683	0.735	38.17	23.57	28.12	34.63	1286.26
Relief	0.646	0.637	0.64	0.631	0.692	77.69	45.22	56.73	61.93	3187.92
IG	0.662	0.641	0.673	0.682	0.726	50.93	25.06	30.28	35.93	1917.64
LTN	0.931	0.923	0.929	0.937	0.946	11.24	8.23	9.3	10.86	438.12



(a) (b)

Figure 8: Continued.

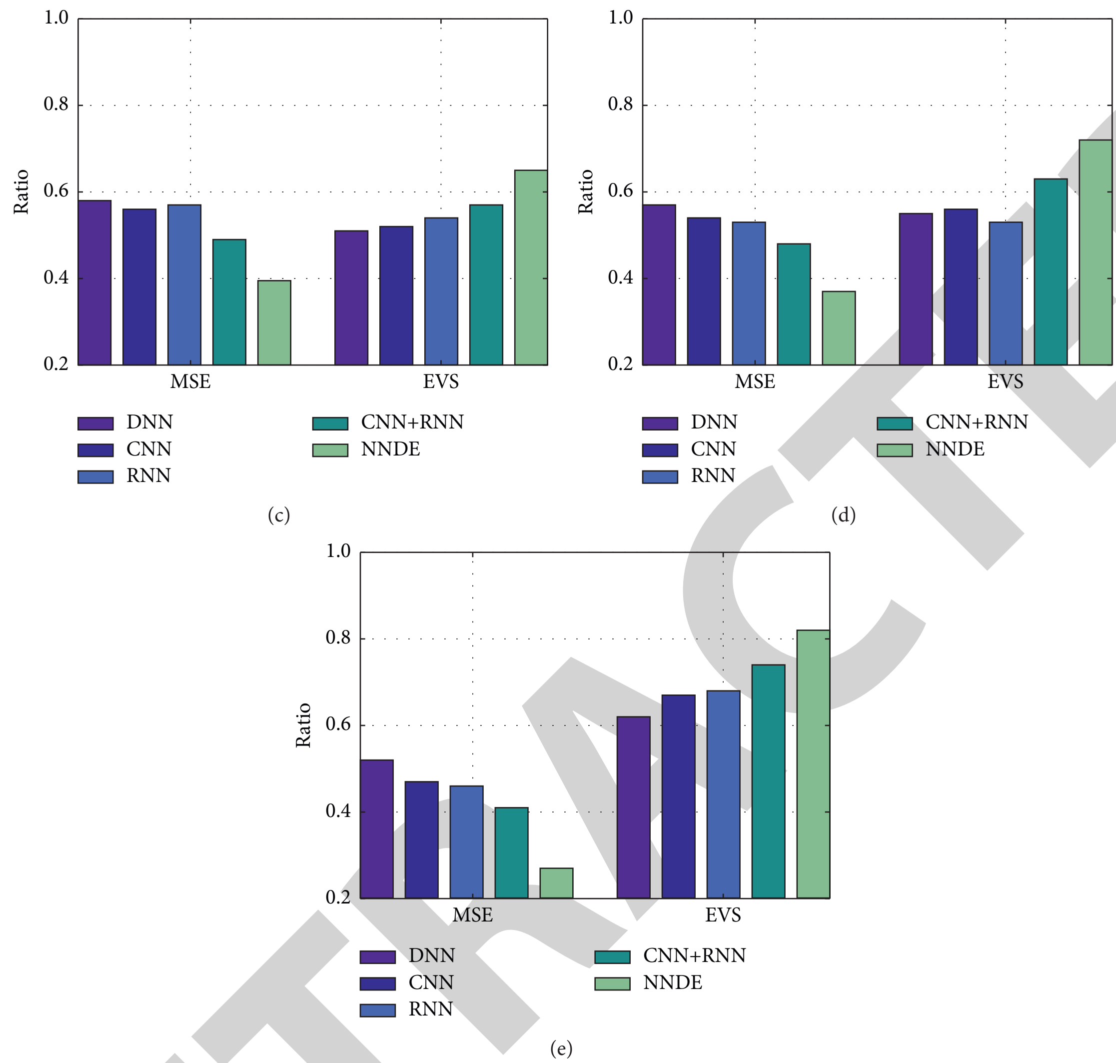


Figure 8: Performance comparison between NNDE and standard neural networks. (a) KDD99, (b) NSL-KDD2009, (c) CIC-IDS2017, (d) CIC-IDS2018, and (e) CIC-DDoS2019.

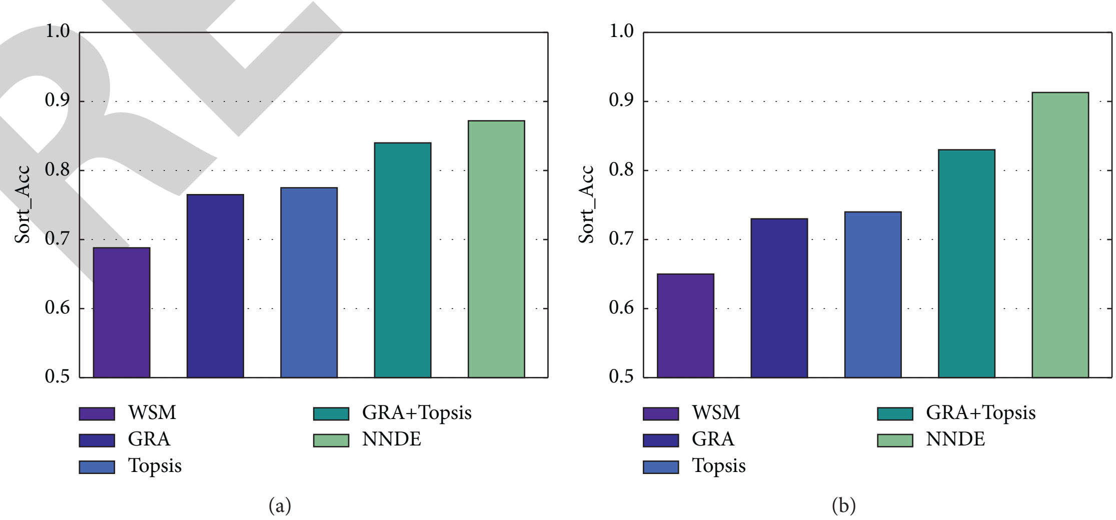


Figure 9: Continued.

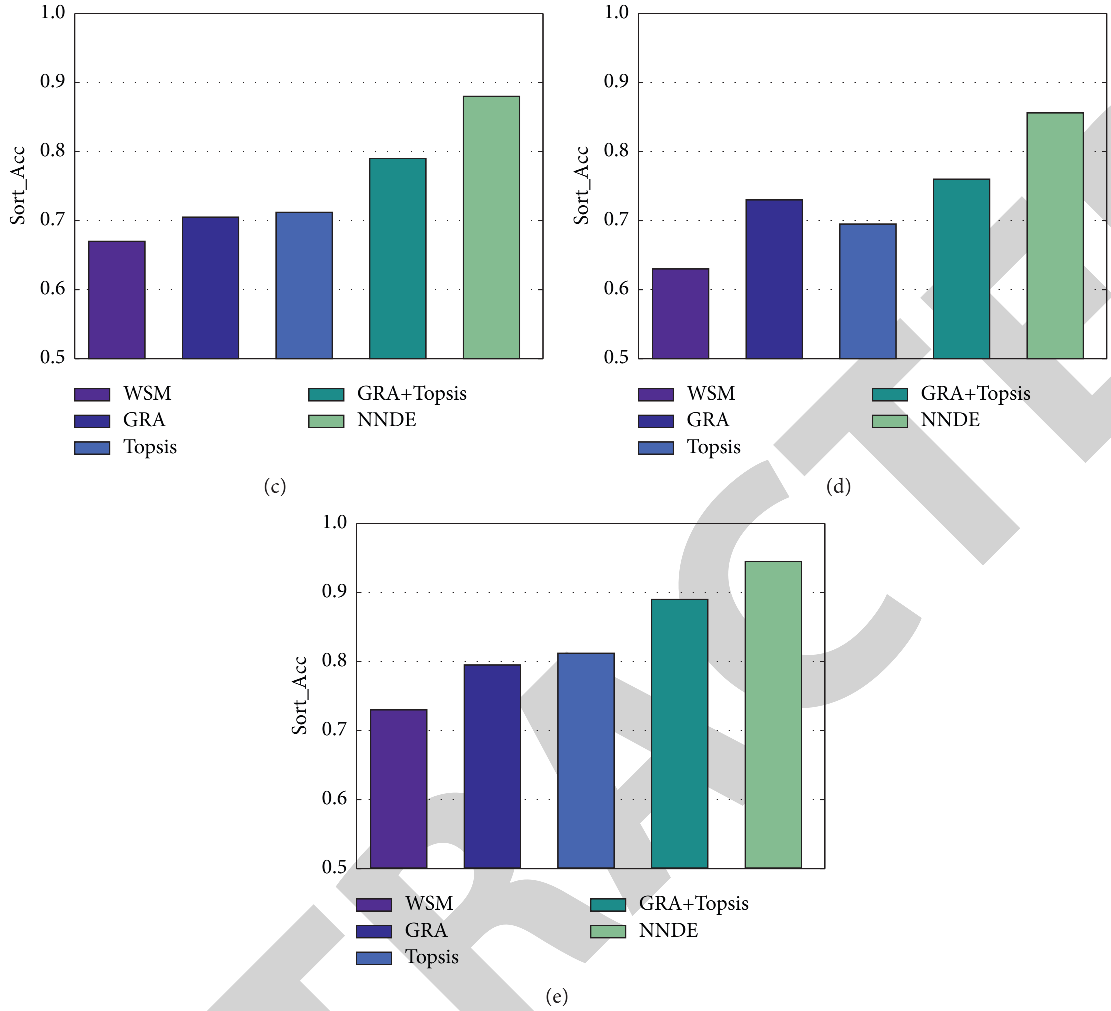


Figure 9: Performance comparison between NNDE and statistical evaluation methods. (a) KDD99, (b) NSL-KDD2009, (c) CIC-IDS2017, (d) CIC-IDS2018, and (e) CIC-DDoS2019.

to reduce the subjective influence. The grey correlation coefficient of GRA is set to 0.5, and TOPSIS is estimated based on the distance from the label to the best and worst solutions. Unlike other neural network structures, in this experiment, Sort_Acc is used to measure the difference between the statistical evaluation method and the NNDE method. The evaluation results on the five datasets are shown in Figure 9. It can be seen from Figure 9 that the NNDE method achieves the highest accuracy rates on all datasets, which are 87.2%, 91.3%, 88%, 85.6%, and 94.5%, respectively. Compared with WSM and combined evaluation methods, the accuracy rates are increased by 21.96% and 7.12%, respectively. Overall, NNDE has achieved an average accuracy improvement of 19.73% compared with other evaluation methods.

Finally, we compared the running time of NNDE with other neural network structures. Each neural network runs ten times, and the average time consumption is shown in Table 5. It can be seen that the more complex the network structure, the more the training parameters, and the longer the running time. Compared with the combined network structure, the time consumption of NNDE has increased by 18.02%. Although the time complexity increases during training, this part of the overhead has no negative impact on our evaluation model application. The reason is that the structure and parameters of the deep learning model are preserved once pretraining is done. Evaluation can be performed using existing models in subsequent evaluations without repeating complex training. Therefore, this added time overhead does not interfere with real-time evaluation.

Table 5: Comparison of running time between NNDE and four neural network structures.

Network structure	DNN	CNN	RNN	CNN + RNN	NNDE
Running time	17.83	21.93	24.92	36.74	43.36

In short, the comparison with four commonly used neural network structures and four classic evaluation algorithms verifies the effectiveness of the proposed NNTE method for DDoS attack effect evaluation. The highest accuracy rate reaches 94.5%. Time expenditure is controlled within a reasonable range to facilitate subsequent defense strategies' rapid deployment.

5. Conclusion

This paper proposes a new method for evaluating the effectiveness of DDoS attacks based on neural networks. We aim to solve the low accuracy of traditional effect evaluation due to the lack of feature selection and targeted evaluation methods. Based on the feature generation tool CICFlowMeter, the redundancy and correlation issues in the DDoS features are analyzed. The relationship between multiple features is captured by combining the TN network and LSTM. Moreover, distance entropy is used to realize the difference measurement of the distribution. In constructing the evaluation neural network, the local features of the vector are captured based on the convolutional layer. The improved attention model achieves hierarchical training of different local features. Finally, feature aggregation is achieved through a fully connected layer. The experimental results show that our proposed neural network-based DDoS attack effect evaluation method could effectively improve evaluation accuracy. More importantly, it provides basic support for the designation of defense plans by security experts.

6. Future Research

We verified the effectiveness of our proposed method from both theoretical and experimental aspects. However, there are still the following problems to be solved.

Question 1. In this article, we generated labels according to equation (1). This method aims at large traffic attacks. Due to similarity with normal traffic, slow DDoS attacks cannot be labeled. So, we need to find a new way to label slow DDoS attacks in the future.

Question 2. In the experiment, we extracted DDoS-related traffic to verify the proposed ideas. Whether this method can apply to other traffic types still needs further verification.

Data Availability

The KDD99, NSL-KDD2009, CIC-IDS2017, CIC-IDS2018, and CIC-DDoS2019 datasets used to support the finding of this study are included within the article.

Conflicts of Interest

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

Acknowledgments

Thanks are due to the artificial intelligence research team of the State Key Laboratory of Mathematical Engineering and Advanced Computing for their help and encouragement. Especially, Dr. Qiu proposed meaningful revisions for the experiments in this article.

References

[1] G. S. Kushwah and V. Ranga, "Voting extreme learning machine based distributed denial of service attack detection in cloud computing," *J. Inf. Secur. Appl.*vol. 53, 2020.

[2] T. Ehrenkranz and J. Li, "On the state of IP spoofing defense," *ACM Transactions on Internet Technology*, vol. 9, no. 2, 2009.

[3] C. Fachkha, E. Bou-Harb, and M. Debbabi, "Inferring distributed reflection denial of service attacks from darknet," *Computer Communications*, vol. 62, 2015.

[4] S. Behal, K. Kumar, and M. Sachdeva, "D-FAC: a novel ϕ-Divergence based distributed DDoS defense system," *J. King Saud Univ. - Comput. Inf. Sci.*vol. 33, 2018.

[5] H. yuan Fang, S. wei Gan, and C. ying Xue, "Evaluation of regional water resources carrying capacity based on binary index method and reduction index method," *Water Sci. Eng.*vol. 12, no. 4, 2019.

[6] W. Wei, L. Zhang, Y. Jiao, C. Tian, C. Wang, and Y. Zhang, "Intracluster structured low-rank matrix analysis method for hyperspectral denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, 2019.

[7] J. R. S. C. Mateo, "Weighted sum method and weighted product method," *Green Energy and Technology*, vol. 83, 2012.

[8] A. Darko, A. P. C. Chan, E. E. Ameyaw, E. K. Owusu, E. Pärn, and D. J. Edwards, "Review of application of analytic hierarchy process (AHP) in construction," *Int. J. Constr. Manag.*vol. 19, no. 5, 2019.

[9] L. Chen, S. Gao, B. Liu, Z. Lu, and Z. Jiang, "FEW-NNN: a fuzzy entropy weighted natural nearest neighbor method for flow-based network traffic attack detection," *China Commun*, vol. 17, no. 5, 2020.

[10] M. Yazdani, C. Kahraman, P. Zarate, and S. C. Onar, "A fuzzy multi attribute decision framework with integration of QFD and grey relational analysis," *Expert Systems with Applications*, vol. 115, 2019.

[11] Y. Qian, X. Liang, G. Lin, Q. Guo, and J. Liang, "Local multigranulation decision-theoretic rough sets," *International Journal of Approximate Reasoning*, vol. 82, 2017.

[12] P. Chen, "Effects of the entropy weight on TOPSIS," *Expert Systems with Applications*, vol. 168, 2021.

[13] M Benassi, "Using two-step cluster Analysis and latent class cluster Analysis to classify the cognitive heterogeneity of cross-diagnostic psychiatric inpatients," *Frontiers in Psychology*, vol. 11, 2020.

[14] H. S. Lallie, K. Debattista, and J. Bal, "A review of attack graph and attack tree visual syntax in cyber security," *Computer Science Review*, vol. 35, 2020.

[15] J. F. Chen, H. N. Hsieh, and Q. H. Do, "Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach," *Applied Soft Computing J.*vol. 28, 2015.

[16] N. Polatidis, M. Pavlidis, and H. Mouratidis, "Cyber-attack path discovery in a dynamic supply chain maritime risk management system," *Computer Standards & Interfaces*, vol. 56, 2018.

[17] D. Gunning and D. W. Aha, "DARPA's explainable artificial intelligence program," *AI Mag*, vol. 40, no. 2, 2019.

[18] L. Liu, W. Ouyang, X. Wang et al., "Deep learning for generic object detection: a survey," *International Journal of Computer Vision*, vol. 128, no. 2, 2020.

[19] M. Sachdeva, K. Kumar, G. Singh, and K. Singh, "'Performance analysis of web service under DDoS attacks," in *Proceedings of the IEEE Int. Adv. Comput. Conf. IACC 2009*, pp. 1002–1007, Patiala, India, February 2015.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9756418, 1 page
https://doi.org/10.1155/2023/9756418

Retraction

Retracted: Hybrid Music Recommendation Algorithm Based on Music Gene and Improved Knowledge Graph

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Zhang and S. Liu, "Hybrid Music Recommendation Algorithm Based on Music Gene and Improved Knowledge Graph," *Security and Communication Networks*, vol. 2022, Article ID 5889724, 11 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5889724, 11 pages
https://doi.org/10.1155/2022/5889724

Research Article

Hybrid Music Recommendation Algorithm Based on Music Gene and Improved Knowledge Graph

Tingting Zhang and Shengnan Liu

College of Music, Shijiazhuang University, Hebei 050035, China

Correspondence should be addressed to Tingting Zhang; 1101838@sjzc.edu.cn

Received 16 February 2022; Revised 10 March 2022; Accepted 21 March 2022; Published 9 April 2022

Academic Editor: Chin-Ling Chen

Combining music as a specific recommendation object, a hybrid recommendation algorithm based on music genes and improved knowledge graph is proposed for the traditional single recommendation algorithm that cannot effectively solve the accuracy problem in music recommendation. The algorithm first gives the recommendation pattern of music genes and gets the relevant recommendation results through the genetic preference analysis. After that, the algorithm in this paper utilizes item and user label information and knowledge graphs from two different domains to enrich and mine the potential information of users and items. In addition, deep learning method is applied to extract low-dimensional, abstract deep semantic features of users and items, based on which, score prediction is performed. The mixed-mode based recommendation addresses the drawbacks of these two recommendations and can adopt different weighting strategies in different situations. The advantages of music gene and knowledge graph-based recommendation algorithms are combined via this method. The experimental results indicate that the algorithm in this paper outperforms other existing recommendation algorithms.

1. Introduction

With the rapid development of mobile Internet, smart terminals, and Internet of Things technologies, a wide variety of information such as text, audio, video, images, and social networks are growing in an explosive situation on the Internet, enriching people's daily life and learning and working content [1, 2]. The rapid development of information technology and Internet technology has generated a huge amount of information, which has greatly not only enriched people's personal needs but also brought about the problem of information overload. Therefore, recommendation systems have emerged, aiming to better meet users' personalized needs and solve the information overload problem. In this era of information overload, the Internet has become the most important part of people's lives. However, a challenge to filter out the interesting contents from the huge amount of Internet information takes up. Recommender systems, which personalize and recommend objects to meet users' needs based on their interests and other characteristics, have been widely used in e-commerce, information portals, social networks, mobile location services, multimedia entertainment, and other fields [3, 4].

Music is an important entertainment element in people's life, and with the development of information technology, music resources are growing in a huge amount [5]. Personalized recommendation, as an effective solution to the information overload problem, has received more extensive attention in the music field [6] and has been widely used. Almost all music platforms currently provide personalized music recommendation services, such as Spotify, Pandara, QQ Music, KuWo Music, and WangYi Cloud Music. Many platforms have achieved a good reputation for the accuracy of recommended music. If only the user and the recommended item are considered, it will affect the performance of the recommendation system to some extent. A very important part of personalized music recommendation system is to consider the user's contextual information at that time and make reasonable use of the user's context for personalized recommendation [7–9]. However, the current music recommendation algorithm is more of a single recommendation algorithm.

Due to the characteristics of music with rich variety, large quantity, short listening time as well as coherence and sequence, the traditional single recommendation algorithm does not address the accuracy of music recommendation in a targeted way. For example, Last.fm is recommended by collaborative filtering, and Pandora is recommended by content similarity [10]. Wang proposed a hybrid recommendation algorithm based on reinforcement learning. The algorithm uses deep learning and weighting matrices to extract song features, considering the problem of changing listener preferences [11]. Feng et al. incorporated attention mechanisms into user history behavior and constructed a hybrid model to finally achieve music recommendation [12], and in both the literature [13, 14], music genes are mixed with recommendation algorithms to achieve music recommendation.

More and more data in the Internet can be accessed sensitively, and multisource heterogeneous auxiliary information including tags, images, and text can be used to enrich the description of users' personalized needs and items and enhance the mining capability of recommendation algorithms. This effectively alleviates problems, such as data sparsity and cold start, and improves the accuracy, diversity, and interpretability of recommendation algorithms. Tags are keywords with no hierarchical structure and employed to describe information, which can be used to describe the semantics of items [15]. Many web-based information service systems, such as product recommendation systems Taobao, Weibo, and Douban, allow users to add tags to items to facilitate information search and item recommendation. Tags serve as a bridge between users and items, expressing the characteristics of items and users' preferences for them. Knowledge graph [16] is a graph-based data structure whose nodes represent entities that exist in the real world and edges represent the relationships between entities. Researchers have constructed a series of knowledge graphs covering different domains, which contain rich information about entities and relationships, extending the hidden relationships that exist between users and items. Knowledge graph can efficiently compute the relationships between entities and provides an effective solution for improving the accuracy and interpretability of recommendation models.

The problem of low accuracy of recommendation algorithms due to existing recommendation systems with problems such as sparse data and cold start is addressed in this paper. In addition, for the problem that the scope of music gene recommendation application is relatively smaller, this paper combines music gene, tagging, knowledge graph, and deep learning, and proposes a hybrid recommendation algorithm based on music gene and improved knowledge graph. The algorithm mixes music gene-based recommendation with improved knowledge graph algorithm by weighting. When the user's evaluation information is relatively small, the weight of music gene-based recommendation is added. When the user's evaluation information is more, the weight of the improved knowledge graph-based recommendation algorithm is increased. This allows deep data mining based on music content features and user preferences, song tags, to improve personalized recommendation efficiency and provide a reference method for optimizing music recommendation systems.

Section 2 in this review is an introduction to related work. Section 3 is on the music gene-based recommendation algorithm. Section 4 is a recommendation algorithm based on improved knowledge graphs. Section 5 is the hybrid recommendation algorithm. Section 6 is the conclusion and Section 7 is the conclusion.

2. Related Jobs

2.1. Tag-Based Recommendation. Tag-based recommendation algorithms are also developing rapidly along with the evolution of tagging technology. Usually, tag-based recommendation systems make recommendations by analyzing the information of tags marked on the items by users. Some researchers use the weight vectors of different tags as input to make recommendations by determining the relevant tags of items and the tag preferences of users. Tag2word models further discover the semantics of tag information to improve recommendation methods. Deep learning is widely adopted in tag-based recommender systems. Some researchers first avail tags to represent user features, and then utilize deep neural network models to extract deep features from the latent space of tags layer by layer and apply the extracted features for collaborative user-based filtering recommendations. Deep-semantic similarity-based personalized recommendation (DSPR) exploits deep semantic similarity-based neural networks to map user features and item features into a deep feature space.

2.2. Knowledge Graph-Based Recommendations. As an effective auxiliary information source in a recommendation system, the knowledge graph contains a large amount of relevant information about the items in the recommendation system and rich implicit semantic associations among the items. RippleNet builds a knowledge graph centered on the items that users interact with, and user interests are propagated outward and decayed layer by layer on the knowledge graph. Knowledge graph convolutional network (KGCN) automatically captures the higher-order structure and semantic information of the knowledge graph using graph attention network and learns the potential long-range interest of users, which makes the recommendation with better interpretability.

In recent years, the recommendation method combining knowledge graph and deep learning has become a development trend, which obtains the low-dimensional vectors of entities and relationships in the knowledge graph through knowledge graph representation learning and uses deep neural networks to abstract the implicit features of items and users to improve the recommendation performance. Deep knowledge-aware network (DKN) extracts the relevant background knowledge of news headlines to construct the knowledge graph, and exploits the TranR method to realize the knowledge. Recurrent knowledge graph embedding

(RKGE) is based on deep neural networks and automatically mines the connection patterns between entities in the knowledge graph to learn the relationship between users and items. Knowledge graph enhanced neural collaborative recommendation (K-NCR) is a neural collaborative recommendation method based on deep neural networks for knowledge graph enhancement, which automatically mines and extends the user's potential interests and connection patterns between entities in the knowledge graph.

3. Gene-Based Recommendation Model

3.1. Gene Structure. Genes are originally biological terms for the basic units of heredity that carry genetic information and control the expression of biological traits. Music includes the fundamental elements of melody, beat, and sound quality. The key to distinguishing one piece of music from another is based on the melody and the beat of the music. The concept of musical genes is also meant to represent the various essential information that controls the auditory effects of music.

Most of the current music sharing systems do not have precise tags about the music, thus users are unable to detect this music through precise music text information. It leads to no small waste of resources, and then the necessity to use music genes to serve users. Generally, if a user likes a certain type of music, it must be because the music has a certain musical characteristic that attracts the user. For example, users' preference for the music of violin playing has a certain relationship with the melodious tone of the violin. Music is rich in content, and in addition to additional attributes (such as composer, era, mood, and other social attributes), music itself exists as a combination of various musical elements (melodies, instruments, and other internal musical attributes). Users can explicitly state the name of the music they like and the type of music they like, but they cannot explicitly state the specific model of the music they like.

If we summarize all the attributes of music, we can basically get the following general structure chart of music genes in Figure 1.

For a piece of music, certain characteristics are fixed genes that exist with the music internally, such as melody, tempo, and other characteristics. These are markers that users can employ to perceive words and emotions, so this category of features can be defined as internal genes of music. As a product born in a society, different music has different social attributes, which are again attributed to social genes. These include the only constant deterministic attributes such as the name of the music, the lyricist, the composer, the name of the singer, the gender of the singer, the language, the region, the genre, etc., and the indeterminate attributes such as emotion, suitability for the situation, etc., which are artificially defined.

Whether users are first exposed to internal genetics (hearing the song first) or social genetics (learning something about the song first) for music genetics, users like a piece of music more in favor of the internal genetics of the music. But the reasons for liking may vary. In most cases, the user simply likes the music, which is a recognition of the music's internal genes. It is also possible that a piece of music that most people do not accept may be changed from not very acceptable to acceptable after repeated exposure because of the user's love for the singer or composer of the music, which is a recognition of the music's social genes.

In the external genetic part of a piece of music, there are some genes that are fixed and will not change, such as the name of the music, the name of the artist, and the gender of the artist. However, there are other genes, such as the emotion and style, which can be different for different listeners. A simple musical genetic structure consists of the following parts:

(1) The name of the music: The name that the author or the producer of the song has drawn up for the song, which is an important marker for listeners when searching for and listening to a song.

(2) Language of the music: The language availed to sing that music, such as Mandarin, English, Cantonese.

(3) The geographical region of the music: the country or region where the artist who sings or plays the music is located, for example, listeners are often classified by Europe, America, Japan, and Korea, the mainland, Hong Kong, and Taiwan.

(4) The emotion of the music that is warm or sad. Even the same piece of music produces different feelings in different people's ears, and listeners can get this free gene by analyzing the labels users put on the music.

(5) The style of music: Works written by certain composers in a certain region of the world or at about the same time often have a similar style, but individual composers using the same musical language can also form individual expressions. Common classifications of musical styles include pop, rock, jazz, punk, etc.

(6) Scenes of different music for different scenes age listening, such as sports time suitable for listening to the passionate music, rest of time suitable for listening to soft and relaxing music.

3.2. Genetic Preference Analysis. In the music recommendation system of this review, three genetic preferences are selected for analysis, namely, geographical preference, emotional preference, and style preference.

3.2.1. Area Preference. By observing listeners' listening behaviors to songs, we can know that different listeners like different music from different regions, such as some listeners prefer European and American music, while some listeners prefer domestic music. For each region, the regional preference of users is calculated separately.

$$AP_{i,j} = \frac{\mathrm{SA}(i, j)}{\mathrm{SC}(i)}, \tag{1}$$

where $\mathrm{AP}_{i,j}$ describes the preference of listener i for songs belonging to region j. $\mathrm{SA}(i, j)$ indicates the number of songs

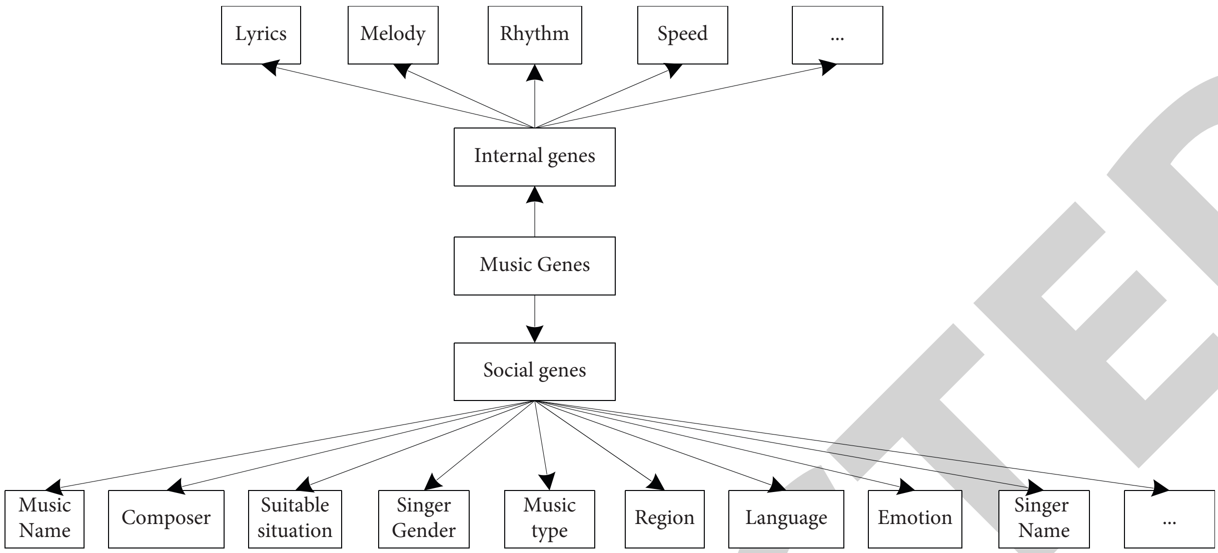


Figure 1: General structure of music gene.

belonging to locale j among all songs listened to by the listener. $\mathrm{SC}(i)$ represents all songs listened to by listener i.

3.2.2. Emotional Preference. Different users have different preferences for music emotions, with some users preferring fresher and warmer ones and others leaning toward slightly sad ones. For each type of music emotion, the user's preference for each emotion is calculated:

$$\mathrm{EP}_{i,j} = \frac{\mathrm{SE}(i,j)}{\mathrm{SC}(i)}. \tag{2}$$

$\mathrm{EP}_{i,j}$ describes how much listener i likes the songs belonging to emotion j. $\mathrm{SC}(i)$ represents all songs listened to by listener i.

3.2.3. Style Preference Degree. Different users have different preferences for music styles, some users like rock punk, others prefer jazz and classical. For each style of music, the user's preference for each style was calculated:

$$\mathrm{SP}_{i,j} = \frac{\mathrm{SS}(i,j)}{\mathrm{SC}(i)}, \tag{3}$$

where $\mathrm{SP}_{i,j}$ indicates the preference of listener i for songs belonging to style j. $\mathrm{SS}(i,j)$ denotes the number of songs belonging to style j in the total songs listened to by listeners. $\mathrm{SC}(i)$ represents all songs listened to by listener i. Finally, the preference of each music attribute above is combined to obtain the genetic preference of the user:

$$P_{i,j} = a^{*}\mathrm{AP}_{i,j} + b^{*}\mathrm{EP}_{i,j} + c^{*}\mathrm{SP}_{i,j}. \tag{4}$$

a, b, and c represent the specific gravity coefficients of the audience to the three attributes of area, emotion and style respectively. The values of a, b and c are set to 0.25, 0.35, and 0.4 respectively, and the sum of the three is 1.

4. Recommendation Algorithm Based on Improved Knowledge Graph

The detailed structure, based on knowledge graph and label-aware recommendation algorithm design is shown in Figure 2, including knowledge graph embedding, user-item modeling, and rating prediction 3parts. The algorithm takes user, item, and label information as input, and obtains feature representations of users and items through the knowledge graph convolutional network to finally predict users' ratings of items.

4.1. Knowledge Graph Embedding. The Knowledge Graph G usually stores entities and their relationships in a triad of "entity-relationship-entity" (b, r, n). The entities and their relations are stored in a triad, where $b \in E, r \in R, n \in E$. E and R are the sets of entities and relations in the knowledge graph, respectively.

4.1.1. Project Entity Embedding. The project embedding reflects the features of project entities. First, DBpedia is applied as a project-oriented knowledge graph to construct a project-centric knowledge graph, and then KGCN is employed to learn the representation of the knowledge graph to obtain the project entity representation vector. KGCN is a graph attention network applied to the knowledge graph, which can be used to capture the topology and the entity information in the knowledge graph. The core idea is to compute the features of a given entity in the knowledge graph by biasedly aggregating and incorporating the information of its neighboring nodes into the features of that entity. The model is a multilayer structure, where the low-order features of an entity can be obtained at the lower level and the higher-order information of an entity can be mined at the higher level.

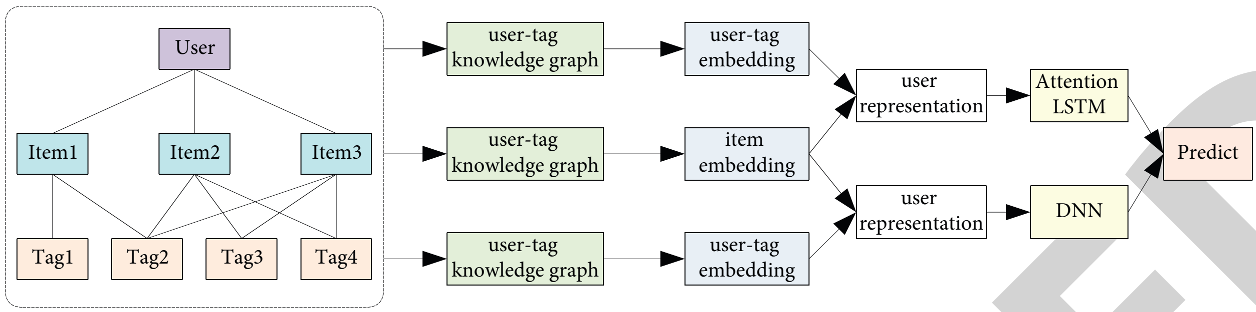


Figure 2: Framework diagram of improved knowledge graph recommendation algorithm.

For users $p \in P$ and items $q \in Q$, where P and Q are the set of users and the set of items, respectively. $T(q)$ denotes the set of entities directly connected to q, and $r_{x,y}$ denotes the relationship connecting entity e_x and entity e_x. The correlation between users and relations is calculated using the vector inner product, which is called the user-relationship score:

$$\pi^p_{r_{x,y}} = p \cdot r_{x,y}, \tag{5}$$

where $p \in \mathbb{R}^d$ and $r_{x,y} \in \mathbb{R}^d$ are the feature representations of the user p and the relation $r_{x,y}$, respectively. d is the feature vector dimension. The user relationship score $\pi^p_{r_{x,y}}$ portrays how important the relationship $r_{x,y}$ is to the user p. For example, a user may have a greater interest in the same type of item. Therefore, a higher attention needs to be given when the relationship $r_{x,y}$ is an item type.

Then the nearest neighbor topology of item q is modeled and the linear combination of domain nodes of item q is calculated:

$$q^p_{T(q)} = \sum_{e \in T(q)} \tilde{\pi}^p_{r_{q,e}} e^0, \tag{6}$$

where $e^0 \in \mathbb{R}^d$ is the initial representation of entity e. $\tilde{\pi}^p_{r_{q,e}}$ is the normalized user-relational score.

$$\tilde{\pi}^p_{r_{q,e}} = \frac{\exp\left(\pi^p_{r_{q,e}}\right)}{\sum_{e \in T(q)} \exp\left(\pi^p_{r_{q,e}}\right)}. \tag{7}$$

In computing the neighborhood representation of an item, the normalized user-relational score reflects the weight of user preferences, and the neighborhood of the item is weighted and aggregated according to the user preference weights.

In a real knowledge graph, the size of neighbors $T(q)$ of different project entities q may vary. To maintain computational efficiency, a fixed number of sets of neighborhoods from the neighbors need to be sampled for each entity. The neighborhood representation of the project entity q is denoted as $q^1_{S(q)}$:

$$S(q) \triangleq \{e \mid e \sim T(q)\}, \tag{8}$$

$|S(q)| = K$ is the number of neighbor samples.

The initial representation q^0 of the project entity and its neighborhood representation $q^1_{S(q)}$ are aggregated to obtain the project entity representation q_{gaa} that incorporates the project entity neighborhood information:

$$q_{gaa} = \sigma\left(M_{gaa} \cdot \left(q^0 + q^1_{S(q)}\right) + h_{gaa}\right), \tag{9}$$

where $M_{gaa} \in \mathbb{R}^{d \times 2d}$ and $h_{gaa} \in \mathbb{R}^d$ are the weights and biases, respectively. σ is the activation function.

With a single-layer KGCN, the representation of an entity will depend on itself and its neighbors, and call q_{gaa} the first-order representation of item q, denoted as q^1. To explore the potential interest of users in a deeper and reasonable way and to explore the higher-order features of entities, the KGCN is extended from one layer to multiple layers. The initial representation of each item entity (i.e., the zero-order entity representation) is propagated outward and aggregated with the representations of its neighboring entities, and then the first-order item entity representation is obtained. The above process is then repeated, that is, the first-order representation is further aggregated to obtain the second-order representation. In general, the L-order representation of an entity is the aggregation of the entity itself with its neighboring entities in the L-hop range. The L-order representation of the item can be taken as the final item entity representation q_e.

First, the knowledge graph is constructed with the item as the center. Second, the neighborhood of the item is sampled and the user-relational score $\pi^p_{r_{x,y}}$, in the graph3 e0i is the initial representation of the entity. Then the first-order representation of the entity e^0_x is calculated by (6). Finally, the first-order representation is extended to the L level, and the L-order representation of the entity e^L_x can be obtained and used as the final representation of the entity.

4.1.2. Label Embedding. The label information is enhanced by using ConceptNet, a word-oriented knowledge graph, to extend the semantic information of labels, aggregate label features with item or user features, and mine the deeper potential features of user and item-related label attributes. The entities in the tags are first identified, disambiguated using entity linking techniques, linked with entities in the ConceptNet knowledge graph, and their associated triples are extracted. Then similar project entity embedding method is exploited to construct the knowledge graph-centered on the entities in the labels, and KGCN method is applied to obtain the representation of the project label q_n and the representation of the user label p_n.

4.2. Project and User Modeling

4.2.1. Project Modeling. Aggregate the project entity embedding q_e with the project label embedding q_n to obtain the final project representation q:

$$q = q_e \oplus q_n, \tag{10}$$

where ⊕ denotes a tandem operation between two vectors. The obtained item characteristics can better reveal the similarity between items.

After obtaining the item embedding, a deep neural network is used to learn the latent features of the item, taking the original item representation $q^{(1)}$ as input and outputting the hidden representation of the item $\widetilde{X}_1$ through an implicit layer:

$$\widetilde{X}_1 = a\left(f\left(q^{(1)}\right)\right) = \mathrm{ReLU}\left(M_1 \cdot q^{(1)} + h_1\right), \tag{11}$$

where M_1 is the weight matrix connecting the input layer to the first implicit layer. h_1 is the bias, and a is the ReLU activation function. In a similar way, the output of the y th layer can be obtained:

$$\widetilde{X} = a\left(f\left(q^{(y-1)}\right)\right) = \mathrm{ReLU}\left(M_p \cdot q^{(y-1)} + h_p\right), \tag{12}$$

where $q^{(y-1)}$ is the output of the $(y-1)$ layer and $\widetilde{X}$ is the potential vector representation of the item.

4.2.2. User Modeling. The user features are aggregated by the item entity embedding q_e and the user label embedding p_n that the user has interacted with according to the temporal organization to obtain the final user representation p:

$$p = q_e \oplus p_n. \tag{13}$$

User characteristics capture both dynamic user preferences for item entities and user preferences for labeled attributes, enabling a better representation of potential user characteristics.

In order to better explore the potential interests of users in the long and short term, a neural network based on the attention LSTM structure is designed, which contains two LSTM layers and an attention layer. The overall structure of this model is shown in Figure 3.

LSTM is a powerful model for learning features from sequence data. Compared with ordinary neural networks, this model performs stepwise analysis of sequences and can model long-range dependencies more effectively by adding hidden layer units that preserve long-term states and storing information from all previous steps in the hidden layer.

The attention mechanism can be used to learn how much the user's historical data contribute to the prediction goal, and thus can effectively reflect the user's behavioral preferences. The attention mechanism calculates the weights of the items that the user has interacted with, and the dynamic preference representation of the user can be obtained by weighting and summing the interaction history representations with these weights. Each item of the vector in the user interaction history $\{q_{e,1}, q_{e,2}, \ldots, q_{e,Z}\}$ is assigned weights

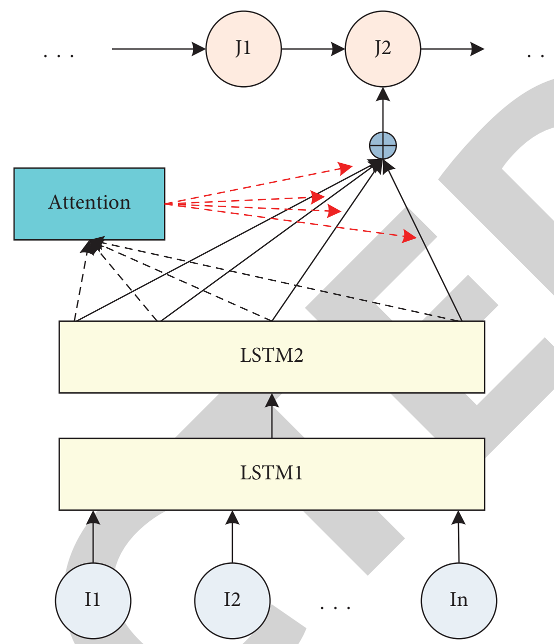


Figure 3: The structure of attention LSTM model.

and weighted equally to obtain the dynamic preference representation of the user.

$$\widetilde{P} = \sum_{x=1}^{Z} g_x q_{e,x}, \tag{14}$$

where, g_x is the attention weight, which is calculated as follows:

$$\begin{aligned} a_x &= \tanh\left(M_{gnn} \cdot q_{e,x} + h_{gnn}\right), \\ g_x &= \frac{\exp\left(a_x\right)}{\sum_{y=1}^{Z} \exp\left(a_x\right)}, \end{aligned} \tag{15}$$

where, M_{gnn} and h_{gnn} are the weights and biases of the attention mechanism.

4.3. Scoring Prediction. The learned latent features of users and items are used to predict the corresponding ratings. In the matrix decomposition method, the inner product of the user potential vector p_x and the item potential vector q_y can be approximated by r_{xy}:

$$r_{xy} \approx p_x \cdot q_y. \tag{16}$$

Therefore, the inner product of two potential feature vectors learned from the hybrid deep structure can be used to obtain the prediction of users' ratings of items:

$$\widehat{r}_{px} = \mathrm{pred}\left(\widetilde{P}, \widetilde{X}\right) = \widetilde{P} \cdot \widetilde{X}. \tag{17}$$

The loss function is the squared error between the true and predicted scores and is the minimization objective of the algorithm, defined as follows:

$$L(\theta) = \sum_{r_{px} \in R} \left(\hat{r}_{px} - r_{px}\right) + \lambda \|\Theta\|^2, \tag{18}$$

where, r_{px} is the actual score. The hyperparameter λ prevents overfitting, which is achieved by controlling the strength of L_2 regularization, and Θ is the set of parameters to be trained. The stochastic gradient descent method is used to minimize the loss function. This algorithm is the basic optimization algorithm in optimization theory, which first finds the direction of the fastest descent by finding the partial derivatives of the parameters, and then continuously optimizes the parameters by iterative method.

5. Mix Recommendation

The main hybrid methods used in current recommender systems can be divided into the following types:

(1) Weighted. By weighting the results of multiple recommendation algorithms to form a final list of recommendations.

(2) Switching.

(3) Mixed.

(4) Waterfall (cascade). By using the latter recommendation algorithm to improve the results obtained by the previous recommendation algorithm. This is a two-step process, the first step uses a recommendation algorithm, the results obtained in this step are relatively coarse, on top of this result, the next step uses another recommendation algorithm to make more accurate recommendations on the results generated by the previous step.

In the hybrid recommendation model of this study, a weighted hybrid recommendation approach is used, that is, collaborative filtering-based and music gene-based, each generates a recommendation result, which is then weighted and mixed to obtain the final recommendation list.

The overall flow of hybrid recommendation is shown in Figure 4.

6. Experiments and Analysis of Results

6.1. Experimental Data Set and Evaluation Index. Dataset1: Last.fm is the world's largest social networking site for music. Users from all over the world listen to songs on this site and communicate with other users. Last.fm also records the songs that each user has listened to. In this experiment, we downloaded the Last.fm dataset and selected a part of it as the experimental data. The selected dataset consists of 2,315 of 29,413 listeners, 29,413 songs, and 16,384 tags.

Dataset2: QQ Music is the most used music software in China, and the software commonly used rate is up to 54.3%. The music dataset2 was obtained by extracting the software data using a crawler tool. The data set contained 3,522 one listener, 35,343 songs, and 10,429 tags.

In the experiments, for each listener, all the music in the dataset is divided into three components A, B, and C, where A is used as the training set and B is used as the test set, with A and B accounting for 8 : 2 of the music listened to by the listener, respectively. c denotes the set of music that the listener has not been exposed to.

In this review, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are employed as evaluation criteria, abbreviated as MAE and RMSE, respectively. MAE is used to measure the quality of recommendation results, and RMSE is availed to measure the variance of recommendation results. The lower the value of MAE, the higher the quality of the recommendation. RMSE is the sum of squares of the differences between the predicted and true values, and can be applied to evaluate the volatility of the predicted results.

6.2. Experimental Results. To evaluate the performance of the algorithm when different network types and different network depths are chosen, a bottom-up approach is used to create different models for comparison. In the algorithm, λ is set as the 0.01, learning rate as the 0.01, and embedding dimension as 100 dimensions. The optimal parameters of the three-layer neural network is determined by experiments on the dataset that are set.

The model1 uses three-layer DNN and single-layer LSTM to learn the latent features of items and users, respectively.

The model2 uses three-layer DNN and two-layer LSTM to learn the latent features of items and users, respectively.

The model learns the latent features of items and users using a three-layer DNN and a two-layer LSTM incorporating an attention mechanism, respectively.

The impact of different structures on the algorithm is shown in Table 1.

From the experimental results, the MAE and RMSE values of the model2 using DNN and two-layer LSTM are 1 lower than those of the models, indicating that the two-layer LSTM structure will have a positive impact on the algorithm performance. This is because the lower layer neural network captures the surface features of the input data, while the deeper layer neural network can extract the higher-level semantic abstraction and obtain the implicit feature representation of the data. Based on the two-layer LSTM structure, the model with the attention mechanism is better 3 than the model2 performance, indicating that the attention mechanism can learn the user's preference weights for items from the user's interaction data and reflect the user's interest, thus improving the recommendation performance. The proposed two-layer LSTM structure with attention mechanism can learn the complex nonlinear relationships in the data through the deep interaction of potential features, and shows better recommendation performance.

Knowledge graph representation learning is to embed the knowledge graph in a low-dimensional space, and the recommendation effect achieved by different embedding dimensions will be different. For the dimensionality of item entity and label embedding, the experiments are conducted in 50–250 dimensions based on the model3, and the experimental results are shown in Figure 5.

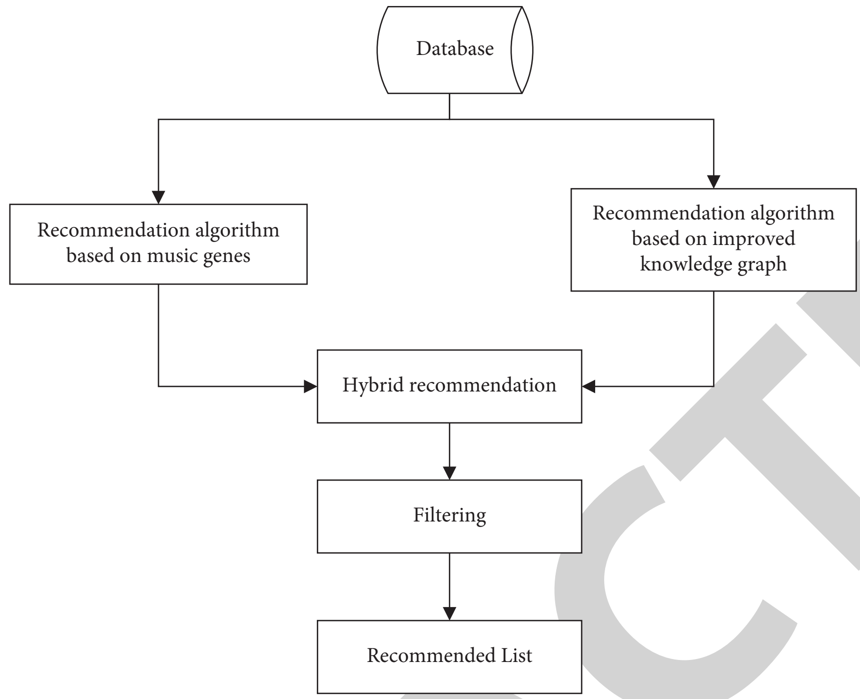


FIGURE 4: The overall flow chart of hybrid recommendation.

TABLE 1: Effect of 1different structures on hybrid recommendation algorithm.

Model	Dataset1		Dataset2	
	MAE	RMSE	MAE	RMSE
Model1	0.834	1.075	0.684	0.891
Model2	0.789	0.986	0.668	0.875
Model3	0.723	0.962	0.612	0.864

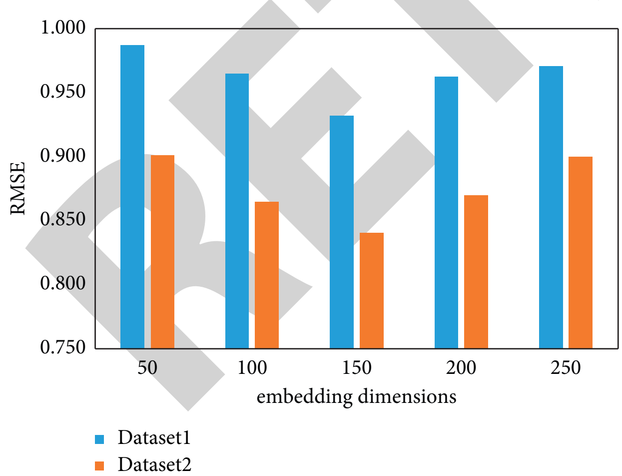


FIGURE 5: RMSE of the algorithm with different embedding dimensions.

From the experimental results, it can be seen that the root mean square error (RMSE) of the algorithm in this study first shows a decreasing trend as the embedding dimension of the knowledge graph representation learning increases. The RMSE value is the lowest when the embedding dimension is 150, and then increases as the embedding dimension increases. The increase of embedding dimension of the algorithm can better express the deep features between users and items, and the model can be better fitted and the recommendation effect is better. However, when the embedding dimension increases to a certain level, the features of users and items in the training set are too detailed, which makes the algorithm overfitting and the prediction effect becomes poor. Therefore, when the embedding dimension is 150, the recommendation effect is best.

In order to verify the effectiveness of the algorithm in this review, it is experimentally compared with five kinds of algorithms, including the label-aware recommendation algorithm and the recommendation algorithm combined with the knowledge graph. The descriptions of the compared algorithms are as follows:

(1) DSPR: Using label-based user and item information as the input of two neural networks, we capture the relevance of users and items using deep neural networks.
(2) TRSDL: Using pretrained word embedding to represent user labels, potential features of items and

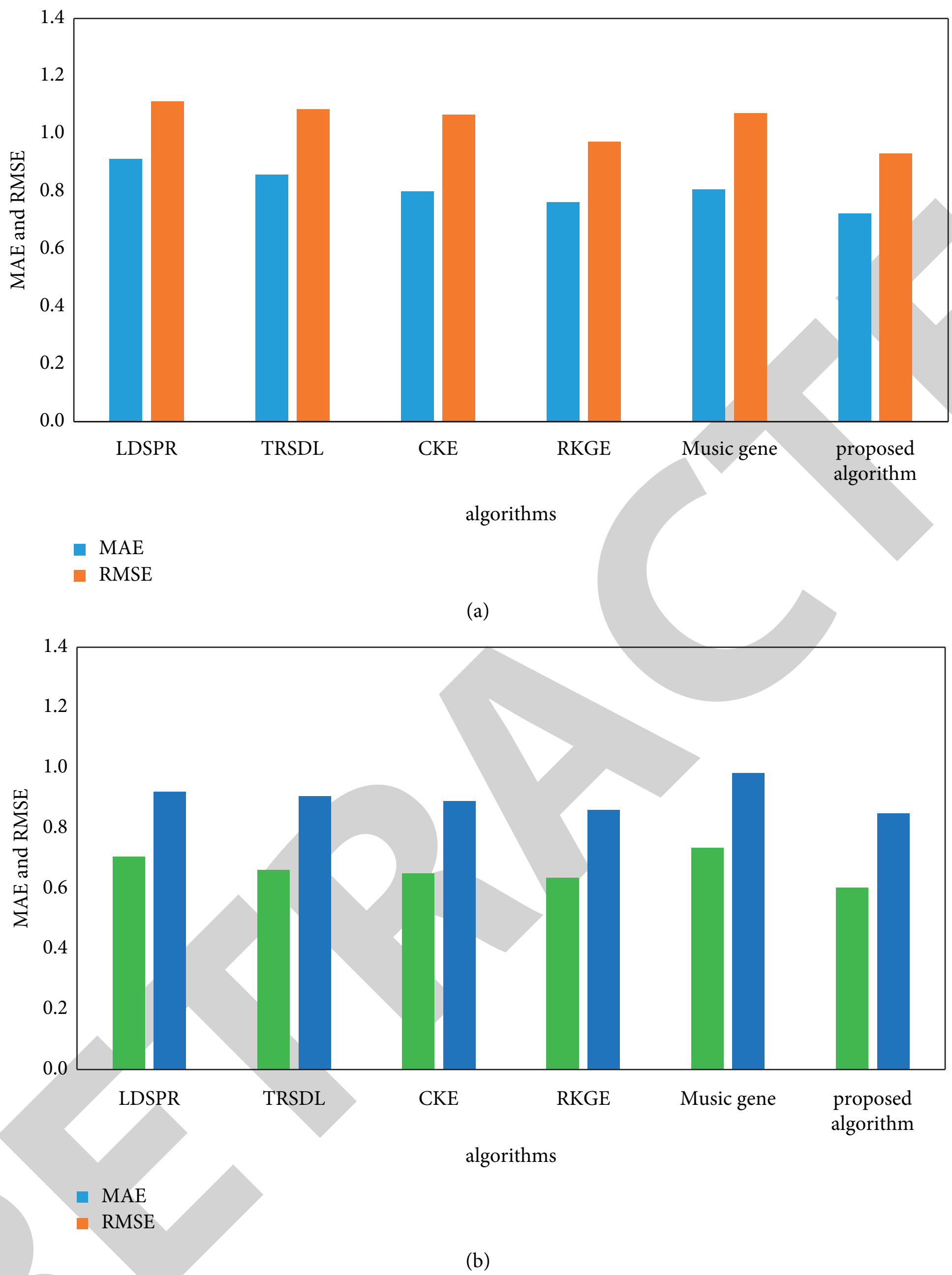


Figure 6: Experimental results of each algorithm on different datasets: (a) Experimental results on dataset1. (b) Experimental results on dataset2.

users are extracted for prediction using DNN and RNN, respectively.

(3) CKE: using a collaborative filtering based approach combined with knowledge graph embedding, items and user representations are learned through a single knowledge graph.

(4) RKGE: using heterogeneous information encoded by knowledge graphs, a recursive network structure is used to learn the relationship between users and items.

As can be seen from Figure 6, TRSDL has better prediction performance than DSPR in that the former takes into account the time factor in extracting label-based user latent features and captures the long- and short-term latent preferences of users through recurrent neural networks. CKE and RKGE also show better performance, both of which use a single knowledge graph of the user's interaction history to construct a representation of the user, and then calculate the entity in the graph correlations with the items to be recommended, indicating that these models can use the knowledge graph to enhance user item information for effective recommendations. Compared with CKE, RGKE takes into account the semantic relationships between path-connected entities and uses recurrent networks to model the semantic paths of entities, thus providing a unified learning method for the representation of entities and entity relationships and improving the recommendation performance. On the other hand, it shows that the CKE method of training entity representations using methods such as TransR do not make good use of the information of the knowledge graph. Recommendation based on music gene does not take user project and user tag information into account, which leads to single recommendation result and small application scope, so the performance of music gene is not good. The hybrid recommendation algorithm proposed in this study outperforms other models on both datasets, which prove the effectiveness of the algorithm. The algorithm in this review combines label information with knowledge graph, uses item-oriented, and word-oriented knowledge graphs to enhance the knowledge of items and labels, respectively, and mines the deeper potential features of users and items through deep structure. At the same time, it can be seen that the two-layer LSTM designed in this review incorporating the attention mechanism plays an active role in the model and improves the prediction accuracy. Finally, the combination with music genes further improves the performance of music recommendation.

7. Conclusion

With the rapid evolution of music business, the music library is getting much richer, and the differentiation of users' preferences is getting much bigger as well. One of the difficulties in music business promotion nowadays is how to make accurate personalized recommendations to users from the huge music library conveniently and quickly. In this study, through analyzing the massive user behavior records kept in music websites, the hybrid recommendation algorithm based on music genes, and improved knowledge graphs utilizes the semantic information inherent in the items themselves and combines user tagging information, which can consider the attributes of items and users more comprehensively. Simultaneously, the algorithm uses two knowledge graphs to enhance the semantic information of items and user tags, and captures low-order and high-order features through a knowledge graph convolutional network. Finally, it combines the music gene recommendation model to make personalized music recommendations for users. The experimental results verify the effectiveness of the algorithm in this study. In future work, more efforts will be made to fuse multi-source heterogeneous auxiliary information and further improve the recommendation performance of the algorithm via extending embedding techniques and deep learning methods.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Shijiazhuang University.

References

[1] R. Kollolu, "Infrastructural constraints of Cloud computing," *International Journal of Management, Technology and Engineering*, vol. 10, pp. 255–260, 2020.

[2] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang, "Big data service architecture: A survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.

[3] P. Nitu, J. Coelho, and P. Madiraju, "Improvising personalized travel recommendation system with recency effects," *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 139–154, 2021.

[4] X. Wen, "Using deep learning approach and IoT architecture to build the intelligent music recommendation system," *Soft Computing*, vol. 25, no. 4, pp. 3087–3096, 2021.

[5] Y. Xu and W. Zhou, "A deep music genres classification model based on CNN with Squeeze & Excitation Block," in *Proceedings of the 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 332–338, IEEE, Auckland, New Zealand, December 2020.

[6] M. S. Fathollahi and F. Razzazi, "Music similarity measurement and recommendation system using convolutional neural networks," *International Journal of Multimedia Information Retrieval*, vol. 10, no. 1, pp. 43–53, 2021.

[7] D. Xu, "Research on music culture personalized recommendation based on factor decomposition machine," *Personal and Ubiquitous Computing*, vol. 24, no. 2, pp. 247–257, 2020.

[8] D. Paul and S. Kundu, "A survey of music recommendation systems with a proposed music recommendation system," *Advances in Intelligent Systems and Computing*, Springer, Singapore, pp. 279–285, 2020.

[9] M. Pulis and J. Bajada, "Siamese neural networks for content-based cold-start music recommendation," in *Proceedings of the Fifteenth ACM Conference on Recommender Systems*, pp. 719–723, ACM, Amsterdam, The Netherlands, October 2021.

[10] D. Wu, "Music personalized recommendation system based on hybrid filtration," in *Proceedings of the 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 430–433, IEEE, Changsha, China, January 2019.

[11] Y. Wang, "A hybrid recommendation for music based on reinforcement learning," in *Proceedings of the Pacific-Asia*

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9821518, 1 page
https://doi.org/10.1155/2023/9821518

Retraction

Retracted: Research on Intelligent Recommendation Model of E-Commerce Commodity Based on Feature Selection and Deep Belief Network

Security and Communication Networks

Received 30 January 2023; Accepted 30 January 2023; Published 2 February 2023

Security and Communication Networks has retracted the article titled "Research on Intelligent Recommendation Model of E-Commerce Commodity Based on Feature Selection and Deep Belief Network" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] Y. Li, G. Wu, and C. Liu, "Research on Intelligent Recommendation Model of E-Commerce Commodity Based on Feature Selection and Deep Belief Network," *Security and Communication Networks*, vol. 2022, Article ID 6469217, 11 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6469217, 11 pages
https://doi.org/10.1155/2022/6469217

WILEY | Hindawi

Research Article

Research on Intelligent Recommendation Model of E-Commerce Commodity Based on Feature Selection and Deep Belief Network

Yunquan Li ,[1] Gaofeng Wu,[1] and Chaohui Liu[2]

[1]*School of Information Engineering, Jiaozuo Normal College, Jiaozuo 454000, Henan, China*
[2]*College of Intelligent Engineering, Zhengzhou University of Aeronautics, Zhengzhou 450046, China*

Correspondence should be addressed to Yunquan Li; liyunquan317@jzsz.edu.cn

Received 17 February 2022; Revised 7 March 2022; Accepted 14 March 2022; Published 7 April 2022

Academic Editor: Chin-Ling Chen

Due to the complexity and uncertainty of customer demand behavior, it was often difficult to obtain satisfactory recommendation results by using the existing online commodity recommendation systems. Therefore, a network commodity intelligent recommendation model based on feature selection and deep belief network was proposed. Based on the basic structure and function of the existing recommendation systems, this paper expounded the interaction process between customers, e-commerce platforms, enterprises, and the recommendation system. By analyzing the internal relationship between customer demand and commodity recommendation, the relationship model between customer demand and commodity recommendation was established. After analyzing the characteristics of customers' demand for goods, a data mining method was used to classify the characteristics of customers' demand behavior, and a feature selection method based on deep belief network (DBN) was proposed to obtain the main information conducive to commodity recommendation. Finally, an e-commerce commodity recommendation algorithm based on feature selection and deep belief network was proposed. The experimental results showed that the network commodity recommendation model proposed in this paper can not only provide customers with satisfactory recommendation results but also has better performance than other traditional recommendation models. The recommendation model proposed in this paper can support different e-commerce website recommendation systems.

1. Introduction

With the improvement of IT technology and network communication level, the e-commerce industry has developed rapidly. The commodity recommendation system provided on the e-commerce website not only provides convenience for customers but also provides strong support for merchants to promote product sales and improve enterprise benefits. A network commodity recommendation system mainly uses customer demand behavior data to establish the corresponding relationship between customers and commodities to predict the potential demand information of customers for commodities [1]. The recommendation system can provide customers with effective recommendation services for online goods. The service provided by the recommendation system can not only stabilize the existing customer market but also meet customers' more needs for goods. Due to the huge network data processing tasks and many data types involved in e-commerce websites, the accuracy, timeliness, and effectiveness of the recommendation system to provide customers with commodity promotion services have attracted the attention of researchers at home and abroad in recent years.

The commodity recommendation system provided by e-commerce websites generally uses customers' previous demand or transaction records for commodities to recommend as many commodities as possible for customers' reference or purchase [2]. Using the recommendation system and network platform, e-commerce enterprises can fully mine the commodity information to meet the needs of customers and provide customers with the commodities they may need to choose and buy at will, which can not only meet the needs of customers but also make profits for businesses. The recommendation system not only provides customers with the goods they need but also provides customers with valuable commodity information [3]. The recommendation system can effectively analyze the customer's historical demand behavior, construct the customer's demand preference model for goods, and use the model to provide customers with accurate commodity recommendation services. Therefore, the network commodity recommendation system can not only meet the needs of customers for commodities but also provide fast commodity sales for merchants to realize the rapid development of e-commerce industry.

From the demand behavior characteristics of customers for online goods, it is very important to use data mining technology to classify a large amount of data and obtain valuable relevant information. In recent years, scholars at home and abroad have proposed different recommendation models or algorithms. These models mainly filter out the data that can support customers' demand behavior from a large number of sparse data information, so as to meet customers' personalized needs for goods [4]. The network commodity recommendation system mainly extracts effective features from the customer's commodity demand behavior characteristics and associates the customer with the commodity information, which not only saves the time for the customer to search the required commodities but also meets the customer's actual demand for commodities. The main function of the recommendation system is to provide customers with feasible demand information, and the recommendation algorithm determines the performance of the recommendation system to a great extent. A good recommendation algorithm can fully meet the needs and interests of different customers and improve the accuracy of the recommendation system. Through the network platform, the recommendation model can not only promote the required commodity information for customers but also obtain the demand intention of different customers for commodities. It is of great significance for stabilizing the customer market and realizing the win-win situation between customers and merchants.

2. Related Works

The function of most recommendation systems is to use some recommendation algorithms to establish the relationship between customer demand objects and recommendation objects. Therefore, an effective recommendation algorithm is an important part of the recommendation system. The idea of recommendation algorithm is to analyze and predict the collection of goods that customers may need to buy according to the collected customer demand behavior information. Common recommendation algorithms mainly include the following four kinds: content-based recommendation algorithm (CRA), collaborative filtering recommendation algorithm (CFRA), association rule-based recommendation algorithm (ARRA), and hybrid recommendation algorithm (HRA) [5–7]. Among them, the recommendation algorithm based on collaborative filtering is one of the most concerned algorithms in recent years. The algorithm can be divided into user-based collaborative filtering algorithm and model-based collaborative filtering algorithm.

In addition to using the commodity recommendation algorithm based on collaboration and content filtering, some scholars have proposed a knowledge-based recommendation algorithm, which can be applied to the recommendation system to effectively analyze customers' demand preferences and purchase behavior. The knowledge-based recommendation algorithm can recommend goods that meet customers' needs according to customers' demand behavior, so customers can really experience the goods and services they are interested in. In addition, some scholars have proposed a utility-based recommendation algorithm. Compared with the traditional content-based recommendation method, the utility-based recommendation algorithm is better than the traditional recommendation algorithm in prediction accuracy, time cost, and customer satisfaction [7, 8]. From the practical application, the performance of utility-based recommendation algorithm usually depends on the context-related information of recommended commodity.

With the continuous development of e-commerce industry, especially the increasing demand of customers and the amount of commodity information, it is necessary to extract the potential information hidden in big data through relevant methods and use relevant recommendation systems to promote the commodities they need for customers [9]. The traditional recommendation algorithm usually cannot obtain the required feature data from a large amount of information, which makes the recommendation results difficult to meet the needs of customers. In addition, the characteristics of many customer demand behaviors are analyzed, and data mining and analysis methods are used to provide customers with accurate product promotion services. In recent years, due to the great progress of data mining technology and deep learning methods, some scholars began to use deep learning methods to study recommendation systems [10]. Many customer demand behavior characteristics are analyzed, data mining and analysis methods are used to screen the characteristics of customer demand for goods, and then an effective recommendation model was used to provide customers with accurate commodity promotion services.

According to the analysis of the recommendation system provided on the existing e-commerce platform, the relevant recommendation models have their own advantages and disadvantages in realizing different customer needs. Because it is difficult to produce good recommendation ability and effect by using a single recommendation algorithm, how to generate recommendation results according to different types of data processing needs and integrate relevant algorithms in order to achieve the best recommendation effect is one of the problems widely concerned by scholars at home and abroad in the construction of recommendation systems. Aiming at the problem of online commodity recommendation, due to the complex change of customers' demand behavior for online commodities and the heavy workload of data processing, this paper proposes an e-commerce commodity recommendation algorithm based on feature selection and deep belief network.

3. Network Commodity Recommendation Theory

3.1. Common Structure of Recommendation System. With the continuous development of e-commerce industry, the use of online commodity recommendation systems has become one of the main tools for communication between merchants and consumers. The design of recommendation system is mainly based on user demand behavior information and goods and services provided by e-commerce enterprises. According to many customers online demand behavior logs, designers can use data mining and analysis methods to obtain customers' demand preferences for different goods and recommend the required goods for different users. Through the recommendation system service, users can improve the click-through rate of online goods to complete commodity transactions or services. It can not only let customers experience the goods they need to buy in advance but also enable businesses to stabilize potential consumer groups and expand the commodity demand market. In order to meet customers' demand for goods, the recommendation system can not only save customers' purchase time to a great extent but also improve the purchase quality and efficiency of goods. Figure 1 shows the relationship between customer demand and e-commerce commodity information established through the recommendation system.

E-commerce companies provide online product recommendation services to customers, usually using media attention to products and using data mining methods to analyze demand behaviors based on customers' past consumption records, and then make suggestions to customers' potential consumption needs. Online commodity recommendation systems mainly customize customers' demand behavior and provide customers with real-time online commodity recommendation services. The main task of online commodity recommendation systems is to recommend appropriate commodities to customers by using different recommendation models according to the needs of users. The structure of online commodity recommendation systems is usually composed of information collection, data preprocessing and analysis, recommendation model, commodity selection, and recommendation. Figure 2 shows the interaction between customers, e-commerce platforms, enterprises, and the recommendation system.

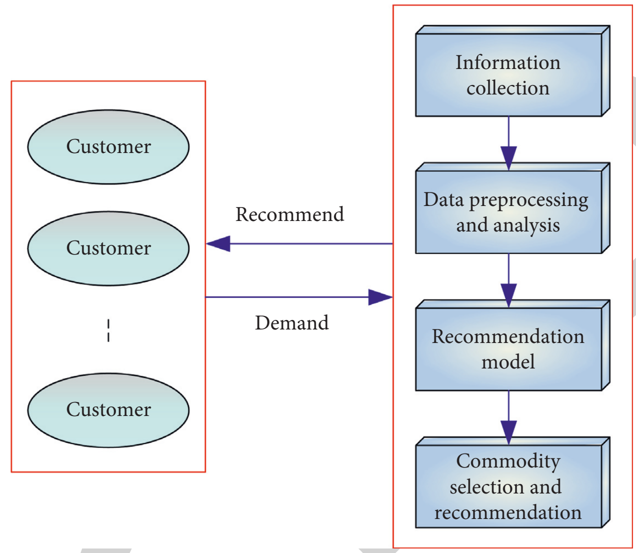


Figure 1: Relationship between customer demand and commodity recommendation.

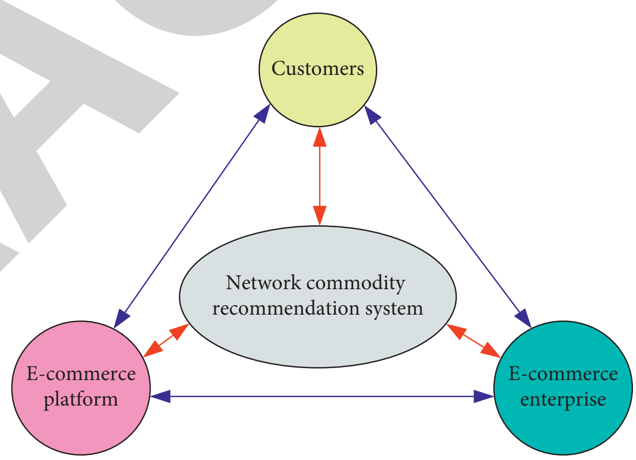


Figure 2: The interaction between customers, e-commerce platforms, enterprises, and the recommendation system.

3.2. Main Functions of Recommendation System. Through the online recommendation system, various networks or mobile platforms can be used to record the merchant website information visited by customers and customers' attention to goods in real time. Based on the data preprocessing of the recommendation system, combined with the user's consumption demand behavior and its associated commodity information warehouse, the original customer demand behavior data set can be constructed. At the same time, the website information or browsing product information visited by the customer is mapped to the recommended candidate product sequence and fed back to the customer as recommendation information [11]. According to the different composition of the online recommendation function, the online recommendation function module can be further divided into customer front-end, background server, backup recommended product collection, and interaction between customers and merchants. The schematic diagram of the personalized online product recommendation process is shown in Figure 3.

Based on the user consumption demand data set obtained in the early stage, the online commodity recommendation system classifies all kinds of commodities by using the method of data mining and forecasts and estimates the user's demand behavior by establishing the customer demand model to realize the online recommendation of different commodities. Firstly, the relevant records of customers and data information such as commodity demand behavior are extracted from the data warehouse, and the irrelevant data are screened and deleted through operations

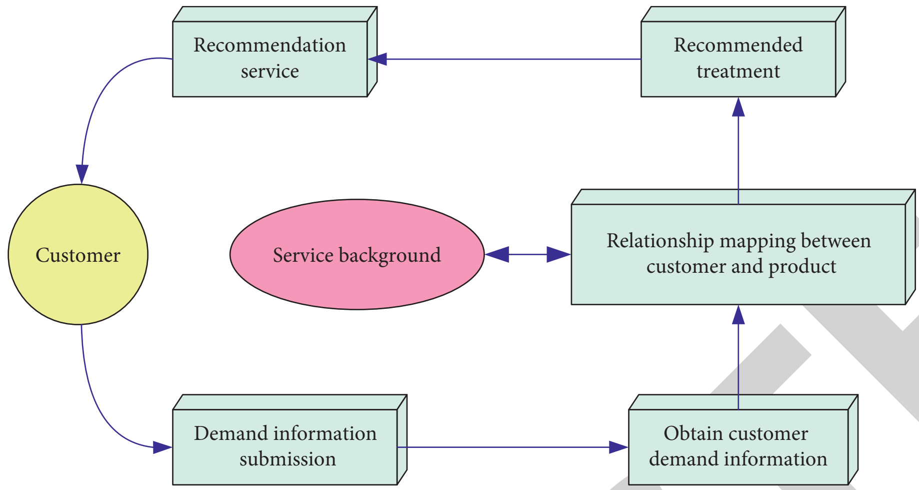


Figure 3: Schematic diagram of personalized online product recommendation process.

such as data preprocessing and feature extraction to retain the required data. Secondly, according to the characteristics of customer demand, the user's demand intention model for goods is established, and the goods required by users are added to the data warehouse. Finally, based on data analysis and in-depth learning methods, the mapping relationship of customers' demand for goods is established, and the goods they intend to buy are recommended to relevant customers.

From the above analysis of online commodity recommendation function, it is known that a typical recommendation system is mainly composed of different customer groups, various e-commerce services, and various website systems including recommendation function. Taking the online product recommendation process as an example, when consumers are willing to demand online products, they need to visit various e-commerce websites containing recommendation functions [12]. E-commerce enterprises generally provide product recommendation services for customers through different e-commerce platforms. The website recommendation system is used to analyze customers' previous demand behavior or purchase transaction records, establish the mapping relationship of customers and corresponding commodities through feature matching, and then recommend appropriate commodities to relevant customers according to customers' demand wishes. A reliable recommendation system should not only provide satisfactory products for customers to choose according to customers' demand behavior but also enable customers to trust the commodity recommendation services provided by the website system. Therefore, a good recommendation system can not only increase the operating revenue for e-commerce enterprises but also effectively meet the actual needs of customers for goods.

3.3. Relationship Model between Customer Demand and Commodity Recommendation. The acquisition of customer demand information is the premise of building a recommendation system model. The acquisition methods of customer demand information data usually adopt active and passive methods. Taking the initiative to obtain customer demand data information generally depends on the active participation of customers, and the participation process may make customers not active enough or even produce boredom and emotion for various reasons. Therefore, this proactive approach is not only easy to lose customers but also difficult to obtain the required customer demand information through big data and artificial intelligence processing methods due to the lack of data. Acquiring customer demand data information in a passive way does not require the direct participation of customers but uses data mining and analysis methods to model the demand behavior of customer history log and commodity transaction information and then put forward the corresponding recommended commodity feature information according to the customer demand behavior data. In recent years, most scholars generally use a passive way to obtain customer demand information when establishing the recommendation system model [13, 14].

In order to build a network commodity recommendation system to meet customer needs, it is necessary to establish a relationship model reflecting customer needs and commodity recommendation, as shown in Figure 4. The relationship model mainly includes customer data information collection and preprocessing, customer demand behavior analysis for goods, establishing the mapping relationship of customer demand for goods, commodity recommendation function, and other modules.

(1) *Collect and preprocess customer data information.* The customer data information to be collected usually includes the customer's purchase history log of goods and various online business activities or transaction behavior information. Based on these data information, we can effectively analyze customers' commodity demand behavior and establish the mapping relationship of customers' commodity demand. Because the collected original data are generally rough and there is lack of correlation between the data, the collected customer's original purchase information can be preprocessed by online

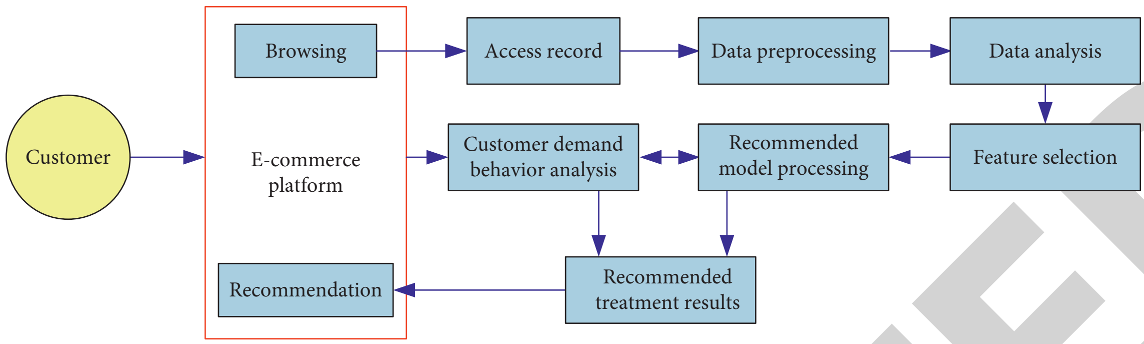


Figure 4: Relationship model reflecting customer needs and commodity recommendation.

detection of transaction records, and the data conducive to completing commodity recommendation for customers can be screened by data mining and feature extraction methods.

(2) *Analyze the customer's demand behavior for commodities*. Use the customer interaction center to obtain the customer's demand behavior data information for commodities, analyze the user's demand for commodities and purchase intention, mine the customer's demand preference for different commodities, and store the associated commodity information in the warehouse center, to provide a data basis for establishing the mapping relationship of customers' demand for commodities.

(3) *Establish the* mapping relationship of *customers' demand for goods*. Generally, by analyzing customers' demand behavior for goods, using deep learning and feature matching methods, combined with the constructed associated commodity information warehouse, customers' demand mapping relationship for different commodities can be established.

(4) *Recommend commodities*. According to the mapping relationship of customers' commodity needs, search for commodities that are greatly related to customers' needs in the related commodity information warehouse, and feedback through the customer interaction platform.

4. E-Commerce Commodity Intelligent Recommendation Model

4.1. Customer Demand Characteristics and Identification. The premise of a good online commodity recommendation system in e-commerce websites is not only to obtain the effective demand information of customers but also to analyze or judge the relevant data to provide the basis for commodity recommendation. For the data information of different objects, it is necessary to mine the potential information hidden in the data through in-depth data analysis [15]. As the recommendation system mainly provides services for customers to select and purchase goods, mastering customers' demand behavior for goods is the prerequisite for building an effective recommendation system. According to the demand behavior characteristics of customers for goods, the relevant information of users and goods can be obtained through data collection and preprocessing. This information reflects the customer's demand for different types of goods or services. Based on the analysis and preprocessing of these data, we can further understand and master customers' demand behavior and purchase intention to provide support for the establishment of an appropriate recommendation system.

Customers' demand or feedback information on goods can be collected through different channels. At present, most e-commerce enterprises mainly obtain it through customers' evaluation of purchased goods and questionnaire survey. This part of information usually reflects the satisfaction of the commodities to the customer according to the customer's objective score or evaluation on the commodities. For example, through the scoring system, customers can score and evaluate the purchase or use of goods, or additional scores can be taken to indicate customers' attitude toward the continuous use of goods. This explicit scoring method can not only obtain the preference of customer groups for goods but also grasp the satisfaction of different consumer groups with different types of goods [16]. Some e-commerce website platforms not only provide customers with scoring mechanisms but also allow customers to comment on commodity needs. No matter using commodity evaluation methods such as comment or scoring mechanism, e-commerce enterprises can timely understand the demand intention of different customers for various commodities, which will provide data support for further obtaining customers' demand behavior.

Affected by the uncertainty of customer demand and the diversification of online commodity types, some feedback information of customers on commodities cannot directly reflect customers' demand intention or satisfaction with the purchase and use of commodities; for example, the number of times customers browse or visit e-commerce websites and customers' attention to businesses or goods. Although this information does not directly reflect customers' demand or purchase intention for goods, it can express customers' various preferences for goods. This kind of information is often referred to as the indirect feedback information of customers to goods. Different from the direct feedback information, although the indirect feedback information has the problems of large amount of data information and difficult to obtain, the data obtained from the indirect feedback information will more truly reflect the customer's demand and preference for goods through data mining and deep learning [17]. From the different behavior characteristics of customers' demand for goods, we can compare the direct feedback information and indirect feedback information from customers, as shown in Table 1.

When analyzing the feedback information of customers on goods, it is usually necessary to analyze the data in combination with the contextual information of customers' demand behavior for goods to deeply understand the changes of customers' demand intention for goods. Generally, different customers have different contextual information about various commodity demand behaviors, and most of the contextual information about customers' online commodity demand behaviors includes customers' access time, place, and mood of e-commerce websites. Therefore, according to the relevant information of customers' demand behavior for online goods, data mining and deep learning methods can be used to extract the time, place, and other characteristics of customers' demand behavior for goods to obtain the information representing customers' demand preference for goods [18].

According to the analysis of customer demand feedback information on different e-commerce websites, it is known that there are certain potential laws in the change of customer demand behavior and online commodity marketing mode. For users, in addition to paying attention to commodity information on e-commerce websites, customers may also pay attention to news and other information. Therefore, in the e-commerce website recommendation system, it is generally difficult for customers to maintain high access frequency and click-through rate for a long time. For online goods, since the newly launched goods are often paid more attention by customers, the visit frequency or click volume of such goods are relatively large at the initial stage of launch, but the attention of goods also decreases with the passage of time. Therefore, when designing e-commerce commodity recommendation system, we should not only consider the changes of customers' demand behavior for commodities but also consider the changing characteristics of commodities and their attributes over time.

4.2. Feature Selection of Customer Demand Behavior Based on Deep Belief Network (DBN). When designing the online commodity recommendation model, there may be some interference data in the customer demand behavior characteristics as the input of the model, which will increase the training time of the model and reduce the efficiency of the recommendation algorithm. Therefore, selecting the characteristics of customer demand behavior is conducive to improve the efficiency and accuracy of the recommendation model [19]. In many customer demand behavior data, the feature processing objects mainly come from customer the demand data set, commodity data set, and the mapping relationship data set between customer and commodity. Because the customer demand data set has the characteristics of discrete distribution, the customer demand data set can be preprocessed by feature reorganization to obtain the required feature set. For the mapping relationship data set between customer and commodity, the data set can be preprocessed according to the customer's demand intention for commodity to obtain the required feature set. In order not to affect the implementation effect of the subsequent recommendation algorithm, because the features extracted from the customer demand behavior are used as the input data of the recommendation model, it is very important to select an appropriate method to extract an effective feature data set.

When completing the task of feature extraction, it is necessary to detect many customer demand behavior data sets, commodity data sets, and mapping relationship data sets between customers and commodities, because the deep belief network model reduces the dimension of the input initial feature set while generating effective features through model training. The deep belief network model contains multilayer neurons, and the bottom layer is the neuron that provides input data for the model, which is called the explicit layer. Other layers except the bottom layer are called hidden layer, which is mainly used to extract data features. Each neuron in the hidden layer represents a kind of features.

Deep belief network is a hybrid model composed of undirected graph model and directed graph model, as shown in Figure 5. Among them, the undirected graph model is an associative memory composed of the top layer and hidden layer of the network, while the directed graph model is composed of restricted Boltzmann machines (RBMs) stacked layer by layer. The features in the directed graph model are trained continuously by the adjacent restricted Boltzmann machine, and finally the selected features are output to the undirected graph model for storage.

The data set of customers' demand behavior for goods includes the operation time of customers' access, click, and transaction, which shows that customers' demand behavior for online goods often changes over time. Therefore, when extracting and selecting the characteristics of customers' demand behavior for goods through the deep belief network model, we need to consider the time context-related information, that is, we need to add the time factor to the feature set based on the deep belief network model. The classification and extraction of different features after model processing are shown in Table 2.

The low-level features of deep belief network can be transferred to the higher-level features after learning and training. Through the continuous learning of features layer by layer in the deep belief network, the high-dimensional complex input features can be trained into low-dimensional features that can not only reflect the information contained in the input data but also clearly reflect the data differences. Therefore, when there are many characteristics of customer demand behavior and the dimension of input data characteristics is high, the use of deep belief network can not only

Table 1: Comparison between direct feedback information and indirect feedback information.

Item	Direct feedback information	Indirect feedback information
Data size	More	Less
Real-time data	Poor real-time performance	Good real-time performance
Customer demand preference	Vague	Clear
Information transparency	Not transparent	Transparent
Information reference value	Necessary	Important

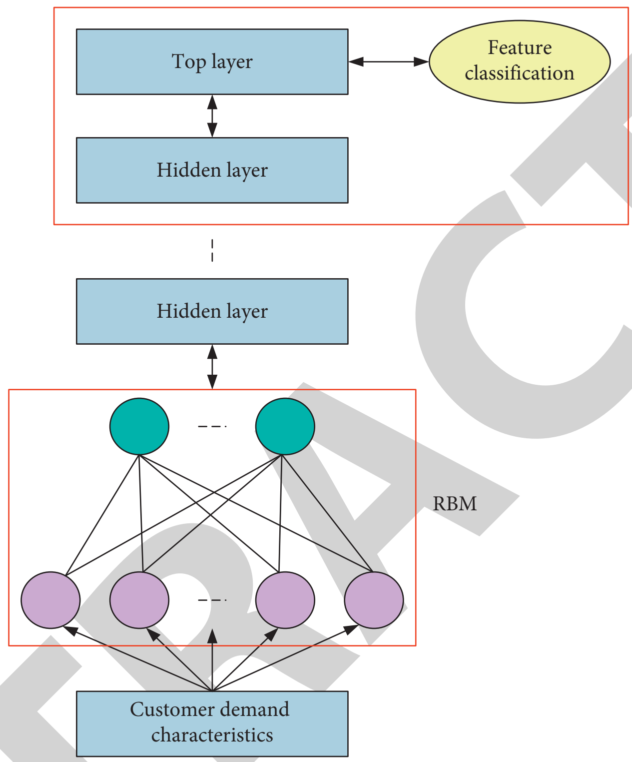


Figure 5: Feature selection process based on deep belief network.

Table 2: Different category features and their attribute extraction.

Feature type	Characteristic content	Feature attribute extraction
*S*1	Customer information	Customer network name, occupation, preference, etc
*S*2	Customer characteristics	Customers click, browse, collect, etc
*S*3	Commodity trading characteristics	Such commodities are clicked, browsed, collected, purchased, etc
*S*4	Characteristics of customers and commodity trading places	Location attributes such as distance between customer and commodity
*S*5	Behavioral characteristics of customer demand for goods	Customer's click, browse, collection, purchase times, and other attributes of goods
*S*6	Commodity characteristics	Commodity click, browse, collection, purchase volume, etc

select the effective features conducive to the recommendation system from many complex features but also improve the efficiency of model training and the accuracy of commodity recommendation.

4.3. E-Commerce Commodity Recommendation Algorithm. According to the design requirements of online commodity recommendation system, based on the construction of customer demand behavior data set, commodity data set, and customer commodity mapping relationship data set, feature detection, extraction, and selection of the customer demand behavior data set based on deep belief network, an e-commerce commodity recommendation algorithm model can be established, as shown in Figure 6.

Firstly, according to the customer demand and the design requirements of the online commodity recommendation system, through the feature extraction and selection of different data sets, combined with the requirements of online commodity recommendation and evaluation, the

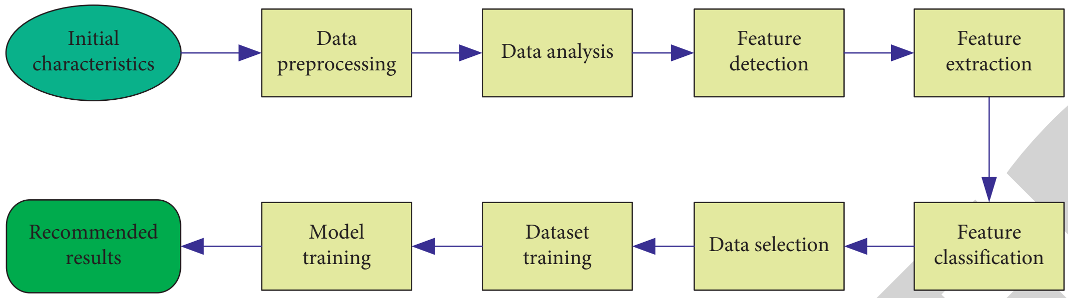


Figure 6: Construction process of e-commerce commodity recommendation algorithm.

original data such as customer demand behavior can be preprocessed [20, 21]. Then, in order to ensure that the feature set can support the recommendation system of the e-commerce platform, some key feature information in the feature set needs to be counted and analyzed in the form of charts. At the same time, in order to improve the accuracy and coverage of online commodity recommendation system and make the recommendation results consistent with customer needs, we also need to consider various factors affecting customer demand behavior. For example, based on the existing data sets, when constructing the recommendation system, we also need to fully consider the information related to context and the information based on the mobile e-commerce platform related to time and geographical location. Therefore, it is necessary to add some additional features to the feature information extracted by the deep belief network. Finally, different data sets are used to strengthen the learning and training of the recommendation algorithm model, and the recommendation algorithm model is optimized by adjusting the model parameters to make the recommendation algorithm consistent with the customer demand behavior.

The performance of online commodity recommendation algorithm largely depends on the processing of customer demand behavior data set, and the results of data processing not only affect the training effect of recommendation model but also directly affect the accuracy and efficiency of the recommendation algorithm. Normal data processing can not only ensure that the algorithm can effectively fit the data set but also mine the internal data relationship in the data set to obtain effective recommendation results. On the contrary, poor data set processing will lead to exceptions in the data trained by the recommended algorithm. For example, the training error of the algorithm means that the customer demand result obtained by the recommendation algorithm is wrong, which affects the accuracy of the recommendation algorithm. Therefore, when constructing the recommendation algorithm model, we need to fully preprocess different data sets to ensure the effectiveness of the recommendation algorithm.

Not all the features in the feature set obtained by feature processing can be used by the recommended algorithm model because, when training the model, these characteristics not only increase the complexity of algorithm implementation but also may affect the implementation effect of the model. Therefore, in the process of feature processing, it is necessary to screen out the features that are conducive to the implementation of commodity recommendation. In order to make the commodity recommendation model meet the needs of different customers, when training the recommendation model with different data sets, it is necessary to screen the feature information that is not conducive to the recommendation results and use the weighted combination method to process the features in different data sets [22]. In order to avoid the overlap between different features, when training different data sets, it is necessary to classify the feature sets and analyze the impact of different types of features on the recommendation model. For those features that can produce better recommendation results for the recommendation algorithm model, the training of these features should be strengthened during model training. When using the deep belief network model to select the features of different data sets, it is necessary to strengthen the learning of training data sets on different feature sets to obtain effective feature sets, which can not only reduce the complexity of the recommendation model but also make the recommendation model obtain better recommendation effect.

5. Results and Analysis

5.1. Model Evaluation Method. In the evaluation of recommendation system, multiple evaluation indexes can be taken from different angles to evaluate the performance of the recommendation system. Different evaluation indicators have different evaluation results on the recommendation system. For example, some indicators describe the performance of the recommendation system in a quantitative way, while some indicators are evaluated in a qualitative way.

The accuracy of the model can reflect the proximity of the recommendation model to the customer demand behavior. This index is mainly used to describe whether the recommendation algorithm can accurately predict the customer demand preference. At present, most recommendation systems use this evaluation index to evaluate the accuracy of the model. In offline calculation of model accuracy, data sets need to be classified differently, including training set, test set, and verification set. When the recommendation algorithm model is used to simulate customer demand behavior, the training set is used to train the recommendation algorithm model, and the test set is used to obtain the recommendation results of the model. Finally, the similarity between the model prediction results and the verification set is used as the accuracy of the model.

In the e-commerce website recommendation system, the commodities recommended to customers according to the law of customer demand behavior are usually given in the form of list. These commodities mainly adopt different recommendation algorithms and are arranged in a certain order. The accuracy of online product recommendation results generated by the recommendation system is mainly reflected by two indicators: recall rate and accuracy rate [7].

The recall rate of the recommendation algorithm can be calculated as follows:

$$\text{recall} = \frac{\sum_{i=1}^{N} f(i) \cap r(i)}{\sum_{i=1}^{N} r(i)}, \tag{1}$$

where N denotes the customer data set, $f(i)$ represents the prediction set of recommendation results obtained by using the recommendation algorithm model trained through the test set, and $r(i)$ shows the actual result set of customer online commodity transaction, that is, the verification set.

At the same time, the accuracy of the recommended algorithm can be calculated as follows:

$$\text{accuracy} = \frac{\sum_{i=1}^{N} f(i) \cap r(i)}{\sum_{i=1}^{N} f(i)}. \tag{2}$$

From the above calculation formula of accuracy and recall, the recall and accuracy used to describe the performance of the recommended algorithm are inversely proportional to each other, that is, the higher the accuracy, the lower the recall. In order to weigh the relationship between the two indicators, a comprehensive trade-off indicator $F1$ can be used to describe the accuracy of online commodity recommendation results of the recommendation algorithm. The calculation formula of $F1$ is as follows:

$$F1 = \frac{2 \times \text{recall} \times \text{accuracy}}{\text{recall} + \text{accuracy}}. \tag{3}$$

Many data feature sets are used to strengthen the training of the recommended model, and the model is optimized by repeatedly adjusting relevant parameters. In order to verify the effectiveness and reliability of the recommended algorithm, the feature test sets of different categories are used as the input values of the model to calculate the corresponding prediction results, and then the feature test sets are used to calculate the recall rate, accuracy rate, $F1$, and other evaluation index values of the model.

In order to judge whether the commodity prediction results obtained by the recommendation system include all commodities that should be recommended, the coverage index can be used to describe. Coverage can better reflect whether the recommendation results cover all recommended commodities to explain the probability of commodities being recommended. Coverage can be expressed by the ratio of the number of all recommended products to the total number of products. Due to the complex demand behavior of customers for goods, in order to meet the various demand behaviors of customers for online goods, the commodity prediction results obtained by using the recommendation algorithm should meet the different needs of customers. Therefore, the recommendation algorithm should usually be diverse. For example, if there is no product that meets the customer's needs in the product prediction result obtained by the recommendation algorithm, the customer is not satisfied with the recommendation result [8]. If the recommendation algorithm can better predict the goods required by customers according to different customer demand behaviors, to meet customers' various needs for online goods, the recommendation algorithm has diversity.

Diversified recommendation algorithms can ensure that customers can choose their own satisfactory products in the product prediction results. For nondiversified recommendation algorithms, there may be products that cannot meet customers' needs in the generated product prediction results, reducing customers' attention and demand interest in online products. Therefore, the diversity of recommendation algorithms can better evaluate the advantages and disadvantages of recommendation algorithms.

5.2. Result Analysis. In order to evaluate the prediction effect of the recommendation model on different characteristics, five different characteristics are selected from the customer demand behavior data set, and the recommendation model is used to predict these different characteristics. According to the different features shown in Table 2, combined with the prediction results of the recommended model, $F1$ values corresponding to these five different characteristics are obtained, as shown in Figure 7.

From the impact of different category features on the recommendation results shown in Figure 7, among the six feature categories, the F1 value obtained when using the recommendation model to predict the $S1$ category features is the largest, indicating that the $S1$ category features have the greatest impact on the recommendation results. The $F1$ value obtained when using the recommendation model to predict the characteristics of $S5$ category is the smallest, indicating that the characteristics of $S5$ category have the least impact on the recommendation results. In the data set of customers' demand behavior for goods, the impact of different categories of features on recommendation results is different, which needs to be determined according to specific different feature categories. Therefore, dividing the customer demand behavior data set into multiple different types of feature sets and using these different types of feature sets to strengthen the training of the model can not only continuously improve the accuracy of the recommendation model but also optimize the effect of the recommendation algorithm.

This paper observed the feature selection process based on the depth belief network model through experiments and analyzed the influence of the number of hidden layer nodes of the model on the recognition accuracy and running time of the model through data statistics, as shown in Figure 8. The experimental results showed that the running time of the model increased linearly with the increase of the number of hidden layer nodes. For the recognition accuracy of the

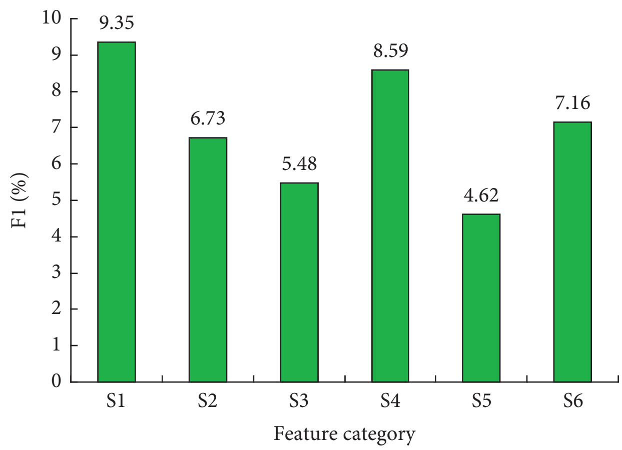


Figure 7: Comparison of accuracy of recommendation results for different types of features.

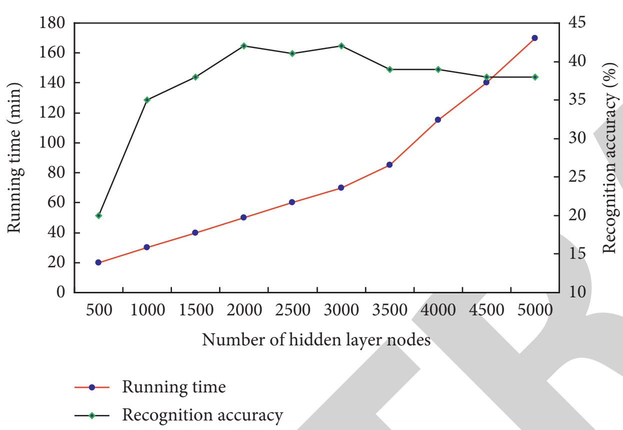


Figure 8: Influence of the number of hidden layer nodes of deep belief network model on recognition accuracy and running time.

model, when the number of hidden layer nodes changed between 500 and 2000, the recognition accuracy showed an upward trend. When the number of hidden layer nodes changed between 2000 and 3000, the recognition accuracy decreased slightly. When the number of hidden layer nodes changed between 3000 and 5000, the recognition accuracy basically did not change.

The performance of different common recommendation models in terms of commodity recommendation time overhead was compared, as shown in Figure 9. The experimental results showed that the recommendation time cost of various models increased with the increase of the number of recommended items, but the recommendation time cost of the model proposed in this paper was lower than that of other relevant models, indicating that the model proposed in this paper had good intelligence for commodity recommendation.

At present, the common recommendation algorithms are mainly collaborative filtering-based recommendation algorithm, association rule-based recommendation algorithm, knowledge-based recommendation algorithm, user behavior analysis-based recommendation algorithm, user

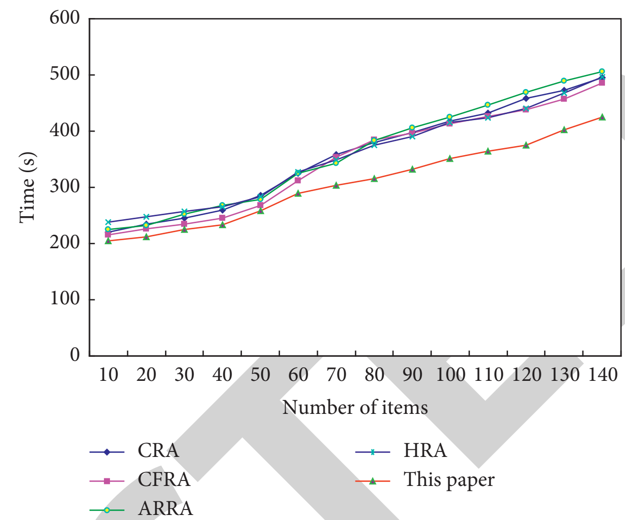


Figure 9: Comparison of recommended time cost of various models.

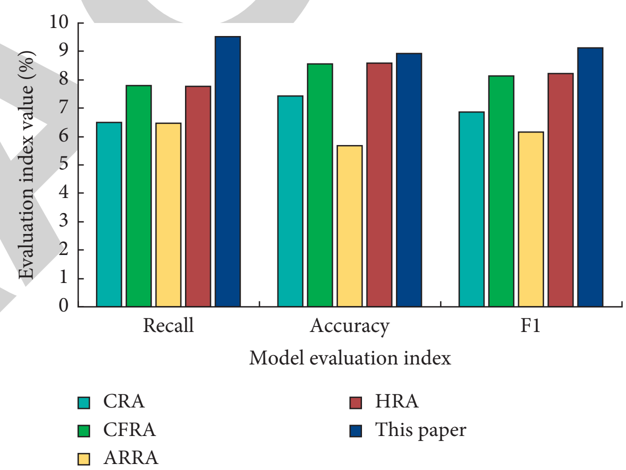


Figure 10: The comparison between this model and other commodity recommendation models.

statistical information-based recommendation algorithm, and utility-based recommendation algorithm. In order to describe the performance of different recommendation algorithms, the recommendation algorithm proposed in this paper are compared with these common recommendation algorithms, as shown in Figure 10. According to the prediction results and actual results of various algorithms, the evaluation index values such as recall, accuracy, and $F1$ of different recommendation models can be calculated. From the comparison results of various algorithms, the recommended algorithm proposed in this paper is superior to other related algorithms in all performance indexes.

6. Conclusion

The traditional recommendation system was difficult to meet the actual needs of customers for goods due to the lack of screening the characteristics of customer demand behavior, and the relevant recommendation models were also difficult

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9845375, 1 page
https://doi.org/10.1155/2022/9845375

Retraction

Retracted: An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 17 November 2022

Security and Communication Networks has retracted the article titled "An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] L. Zhou, Q. Chen, and T. Zhu, "An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations," *Security and Communication Networks*, vol. 2022, Article ID 2173402, 11 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2173402, 11 pages
https://doi.org/10.1155/2022/2173402

Research Article

An Improved Data Mining Model for Predicting the Impact of Economic Fluctuations

Liting Zhou,[1] Qiang Chen,[2] and Tianyu Zhu [3]

[1]*Department of Finance, Huaibei Vocational & Technical College, Huaibei 235000, China*
[2]*Training Centre, Huaibei Vocational & Technical College, Huaibei 235000, China*
[3]*School of Information, Huaibei Normal University, Huaibei 235000, China*

Correspondence should be addressed to Tianyu Zhu; zhuty@chnu.edu.cn

Received 8 January 2022; Revised 27 February 2022; Accepted 11 March 2022; Published 31 March 2022

Academic Editor: Chin-Ling Chen

In order to explore the influencing factors of economic fluctuations, this article combines data mining technology to mine the factors that affect economic fluctuations and introduces financial network theory to the transfer problem of investment strategies, which provides an effective research method for economic behavior research. Moreover, this article calculates the fund network parameters from a quantitative perspective by constructing the topological characteristic index of the financial network to illustrate the influence of the degree of network evolution and the time scale. In addition, this article constructs an economic fluctuation data mining model based on economic fluctuations and uses data to verify the effects of the method proposed in this article. The experimental results show that the data mining method proposed in this article can play an important role in the analysis of factors affecting economic fluctuations and accurately mine the relevant factors that affect the economy.

1. Introduction

With the continuous advancement of economic globalization and world economic integration, countries in the world have formed interconnected and interdependent organisms to a certain extent through economic activities such as trade and investment [1].

In order to promote the rapid recovery of their own economies after the financial crisis, governments around the world have implemented economic policies more actively. However, the frequent promulgation of economic policies has increased the complexity of economic ties between countries, reduced the stability of the country's economic system, and aggravated the uncertainty of economic policies, thereby adversely affecting macroeconomic fundamentals and slowing the speed of economic recovery. There is currently no authoritative definition of the concept of economic policy uncertainty. The mainstream view is that the inconsistency between the current and expected views of economic participants on the impact of economic policies on the economy has caused economic policy uncertainty, and economic uncertainty is closely related to economic operations and economic policies. The variability of the market makes possible mismatches and time lag in the transmission of economic policies, resulting in failure to produce the expected effect of resource allocation by the government, resulting in uncertainty in economic policies. Especially during the financial crisis, while the government implemented a series of fiscal or monetary policies to stimulate the economy, it also greatly increased the uncertainty of economic policies [2].

The current methods for measuring the uncertainty of economic policies can be summarized into three categories according to the measurement ideas: proxy index substitution method, subjective expected deviation method, and state quantity estimation method. The more common in the proxy index substitution method is the frequency of keywords such as economic uncertainty in statistical newspapers and periodicals. This method believes that the formulation and influence of economic policies and people's views on the economy must be reflected in the media. By counting the frequency of occurrence of keywords related to uncertainty, economic uncertainty can be reflected, to a certain extent. However, the choice of newspapers and media depends entirely on the subjective wishes of statisticians, and the views of newspapers and media do not represent the subjective perceptions of all members of the economy. Therefore, it is difficult to fully reflect the macroeconomic conditions [3].

The occurrence of major events has varying degrees of impact on the politics, culture, and economy of various countries. As a core content of world development, the financial market is also greatly affected. The financial market has a core position in the social economy, and the most important part of the financial market is the securities market, so the securities market can represent the financial market to a certain extent. The occurrence of major events will cause volatility in the stock market and affect the correlation between stock markets.

At present, some scholars only use the univariate GARCH model to discuss the volatility of a certain stock market without considering the correlation between the markets; other scholars consider the correlation between the markets and use the multivariate GARCH model to analyze the stock market; and relatively little research has been done on how the occurrence of major events affects the stock market. The influence of the securities market on the development of the country and people's social life cannot be ignored. Therefore, it is of practical significance to study the impact of the occurrence of major events on the correlation between the stock markets, whether it is for the development of society or the investment of investors.

The organizational structure of this article: the first part is the introduction, which mainly analyzes the research background, research status, research motivation, and research significance. The second part is the literature review part, which summarizes and analyzes the quantitative methods of economic fluctuations. The part is to improve the big data algorithm, which provides the basis for the intelligent method of the fourth part. The fourth part is the model construction part of this article, and the fifth part is the system experiment part of this article. The conclusion part is the research content of this article. The summary and outlook for future work are then presented.

The main contributions of this article are as follows: (1) big data methods are used combined with financial network theory to explain the phenomenon of economic fluctuations and dynamically identify the changing process of fund investment strategies. (2) It provides an effective method for the prediction and analysis of economic fluctuations in the era of big data.

This article analyzes the economic fluctuations with big data technology and obtains the main factors that affect economic fluctuations through data mining and the economic fluctuations are analyzed. Moreover, this article combines experimental research to evaluate the effect of economic fluctuation forecasting to verify the reliability of the method.

2. Related Work

Academia usually adopts VAR and GARCH family models to measure spillover effects between markets. Zhou et al. [4] used the recursive VAR method to examine the relationship between quantitative easing in the United States and the volatility spillover effects of major international financial markets. It is found that the unconventional monetary policy of the United States has a significant impact on volatility spillovers and potential global systemic risks. Bhattacharya et al. [5] measured the time-varying spillover effects between my country's real economy and the stock market and bond market by constructing a mixed-frequency VAR model. The study found that the time-varying characteristics of the spillover effect between the real economy and the two markets are significant, and the spillover effect increased during the financial crisis, and then the spillover effect decreased. Geng et al. [6] used the TVP-VAR model to calculate the time-varying volatility spillover index. Vu et al. [7] pointed out that there is an asymmetric two-way volatility spillover between the offshore RMB exchange rate and the onshore exchange rate. Teljeur et al. [8] found that the volatility of the stock index of the sample countries had an enhanced spillover effect on the volatility of the stock index, and after the financial crisis, there was a leverage effect and spillover effect of the volatility of the sample interest rate on the volatility of the stock index, but the impact was minimal. Rajsic et al. [9] found that volatility spillovers between financial markets have significant time-varying characteristics. Jahedpari et al. [10] examined the volatility spillover effects among 21 stock markets in Asia, Europe, Africa, and the Americas. Daksiya et al. [11] established a wavelet multiresolution BEKK-GARCH model for the return sequence of the foreign exchange market and the stock market. It turns out that in the low-frequency domain, there is a one-way volatility spillover effect from the stock market to the foreign exchange market, while in the high-frequency domain, there is a two-way volatility spillover effect between the two markets.

Lahmiri [12] used rolling regression and event research methods to verify the long- and short-term effects of the US quantitative easing policy on gold prices. The results show that quantitative easing policies have a significant impact on gold prices in both the long term and the short term. Gordini [13] used the spillover index and complex network method to measure the intensity and direction of China's financial risk spillovers and found that China's financial risk spillovers have volatility, uncertainty, and asymmetry, and the lagging effect of each market is obvious. Ferramosca et al. [14] studied the volatility spillover effects between international crude oil prices, US economic uncertainty, and Chinese stock markets by constructing static and dynamic volatility spillover indexes. Jane [15] selected the vector autoregressive model and the asymmetric BEKK model to empirically study the spillover effects between China's stock market, foreign exchange market, and currency market. Nassirtoussi et al.

[16] studied the spillover effects among four crude oil markets, including China. Ellis and Christofides [17] used the VAR-BEKK-GARCH model to investigate the volatility spillover effects of China's fuel oil spot and fuel oil futures and energy stock markets. The empirical results show that there is a two-way volatility spillover effect between fuel futures and the spot, while only the energy stock market has a one-way volatility spillover effect on the fuel oil futures market.

3. Principles of Financial Networks Based on Big Data Algorithms

3.1. Analysis of Financial Parameters. $r^{\Delta t}(t)$ is the percentage change of daily net value Z at time t and time scale (span) Δt. In order to eliminate the error caused by the difference in the net value of different funds and maintain the stability of the data series, a first-order logarithmic difference is carried out on the net value of the fund:

$$r^{\Delta t}(t) = \ln Z(t+\Delta t) - \ln Z(t). \tag{1}$$

Based on the logarithmic rate of return $r^{\Delta t}(t)$, the Pearson correlation coefficient $\rho_{ij}^{\Delta t}$ between fund i and fund j on the time scale Δt can be calculated by the following formula :

$$\rho_{ij}^{\Delta t} = \frac{\langle r_i^{\Delta t}, r_j^{\Delta t}\rangle - \langle r_i^{\Delta t}\rangle\langle r_j^{\Delta t}\rangle}{\sqrt{\left(\langle\left(r_i^{\Delta t}\right)^2\rangle - \langle\left(r_j^{\Delta t}\right)^2\rangle\right)\left(\langle\left(r_j^{\Delta t}\right)^2\rangle - \langle r_j^{\Delta t}\rangle^2\right)}}. \tag{2}$$

$\rho_{ij}^{\Delta t}$ represents the correlation between the net value of fund i and fund j on a time scale of Mr. Among them, $\langle\cdots\rangle$ represents the expectation operator. $\rho_{ij} \in [-1, 1]$ represents the correlation coefficient, which measures the degree of correlation between the actual strategies of funds. When $\rho_{ij} = 0$, it means that there is no correlation between the actual strategies of the two funds. When $\rho_{ij} > 0$, the actual strategies of funds i and j are positively correlated, and the larger ρ_{ij} g, the greater the degree of correlation between the actual strategies of the two funds. When $\rho_{ij} = 1$, the two funds adopt the same actual strategy. When $\rho_{ij} < 0$, the actual strategies of funds i and j are negatively correlated, and the larger ρ_{ij}, the smaller the correlation between the actual strategies of the two funds. When $\rho_{ij} = -1$, the two funds adopt opposite actual strategies. Due to the finiteness of the time series of fund net value, there may be false correlations between funds. The correlation coefficient of fund samples is used to infer whether the actual strategies of the two fund variables in the fund population are related. The significance test of the null hypothesis that the overall correlation coefficient is 0 can be carried out using the t-distribution statistics of the overall degree of freedom df $= n-2$ (n is the number of fund samples) for the overall correlation coefficient of the fund (formula (5)). This article will take a hypothesis test on the correlation coefficient at a 95% confidence level. If the t test is significant, formula (3) holds, which means that there is no correlation between the actual strategies of the two tested funds, namely, $\rho_{ij} = 0$. If the t test is not significant, formula (4) holds, indicating that there is a correlation between the actual strategies of the two funds, and the correlation coefficient is calculated by formula (2).

The hypothesis test is as follows [18]:

$$H_0\colon \rho = 0, \tag{3}$$

$$H_1\colon \rho \neq 0, \tag{4}$$

$$t = \rho_{ij}\sqrt{\frac{n-2}{1-\rho_{ij}^2}} \sim t(n-2). \tag{5}$$

The correlation coefficient matrix between funds cannot be directly represented by a network diagram. The distance between funds is calculated by the Euclidean distance formula and the correlation coefficient matrix is converted into a distance matrix to construct an undirected weighted financial network to describe the fund market.

We set the following:

$$\tilde{r}_i = \frac{r_i - \langle r_i\rangle}{\sqrt{\langle r_i\rangle^2 - \langle r_i^2\rangle}}. \tag{6}$$

Among them, $\tilde{r}_i$ is the time series of the net value of fund i. The n records of the vector $\tilde{r}_i$ in the same time interval are taken as the distance $\tilde{r}_{ik}$ between the points of the n-dimensional vector $\tilde{r}_i$. Then, d_{ij} of the two funds can be obtained from the Pythagorean relationship:

$$d_{ij}^2 = \left\|\tilde{r}_i - \tilde{r}_j\right\|^2 = \sum_{k=1}^{n}\left(\tilde{r}_{ik} - \tilde{r}_{jk}\right)^2. \tag{7}$$

From the definition of $\tilde{r}_i$, the length of the vector $\tilde{r}_i$ is 1; that is,

$$\sum_{k=1}^{n}\tilde{r}_{ik}^2 = 1. \tag{8}$$

Therefore, formula (7) can be rewritten as follows:

$$\begin{aligned} d_{ij}^2 &= \sum_{k=1}^{n}\left(\tilde{r}_{ik}^2 - \tilde{r}_{ik}^2 - 2\tilde{r}_{ik}\tilde{r}_{ik}\right)^2 \\ &= 2 - 2\sum_{k=1}^{n}\tilde{r}_{ik}\tilde{r}_{ik} \\ &= 2 - 2\rho_{ij}. \end{aligned} \tag{9}$$

From formula (8), we can get the following:

$$d_{ij} = \sqrt{2\left(1-\rho_{ij}\right)}. \tag{10}$$

Among them, $\tilde{r}_i$ is equivalent to the Euclidean distance of the net value time series vectors of any two funds i and j, $d_{ij} \in [0, 2]$. It is satisfied with the three properties that must be satisfied by the Euclidean distance, namely (1), $d_{ij} = 0 \Leftrightarrow i = j$ (2): $d_{ij} = d_{ji}$ (3): $d_{ij} \leq d_{ik} + d_{kj}$. The properties (1) and (2) are easy to verify because $\rho_{ij} = \rho_{ji} \Leftrightarrow d_{ij} = d_{ji}$. The triangle inequality of property (3) can be proved by the equivalence of (7) and (10). The economic meaning of $\tilde{r}_i$ is the correlation coefficient between funds, and the smaller the distance, the more similar the actual strategies between funds, and vice versa [19].

3.2. Analysis of Financial Network Algorithms. The establishment of a fund network of N funds can be calculated by calculating the correlation coefficients between N funds. Calculate the distance between funds by formula (8). If the hypothesis test of the correlation coefficient between the funds is not considered, N funds will construct a distance matrix of $N \times N$, and a complete graph representing the actual strategy network of the funds can be constructed through the distance matrix between the funds. A complete graph is a simple graph in which each pair of different vertices is connected by an edge. Therefore, the fund strategy network of the complete graph has a large amount of information and is not easy to handle. Therefore, it is necessary to use a specific method to filter some redundant edges while retaining the valuable actual strategy edges in the fund network to form the final fund actual strategy financial network. Next, starting from the fund distance matrix, the final fund actual strategy financial network is generated based on the minimum spanning tree method and the planar maximum filter graph method.

There are two basic algorithms for minimum spanning trees: Kruskal's algorithm and Prim's algorithm. The principle of the Kruskal algorithm mainly starts from the connection edges and gradually determines the connection edges that meet the conditions to obtain the final minimum spanning tree network. Prim's algorithm expands around nodes and determines the nodes and connecting edges in the network one by one to get the final minimum spanning tree network. There is no essential difference between the two algorithms, and the final result is the same. In this article, Kruskal's algorithm is used to screen the edges of the distance matrix step by step to obtain the minimum spanning tree network. The specific methods of Kruskal's algorithm are as follows:

(1) The algorithm arranges the weights of all connected edges (that is, the distance of the fund) in ascending order and selects the edge with the smallest weight:

(2) In each step, the edge with the smallest weight is selected from the unselected edges so that it does not form a circle with the selected edges, until the $N-1$ edges are selected.

The Kruskal algorithm of the minimum spanning tree is programmed by Matlab7.0.

We assume that the space of N funds is a hypermetric space. This hypothesis is based on the "posterior" motivation; that is, the research results obtained based on this hypothesis are meaningful from an economic point of view. The supermetric space refers to the space in which the distance between objects is the supermetric distance. The supermetric distance satisfies the first two properties of distance, namely, (1) $\widehat{d}_{ij} = 0 \Leftrightarrow i = j$ and (2) $\widehat{d}_{ij} = \widehat{d}_{ji}$. However, the property (3) of distance The triangle inequality is replaced by the hypermetric inequality, namely, $\widehat{d}_{ij} \leq \max\{\widehat{d}_{ik}, \widehat{d}_{kj}\}$. Rammal et al.'s article introduced the concepts related to hypermetrics in detail. Several supermetric spaces can be obtained by segmenting a set of N objects with a certain metric distance relationship. Among all possible hypermetric structures corresponding to distance d_{ij}, the subhypermetric space is the simplest and has good properties. In the metric space where N objects are associated together, the subsupermetric space can be obtained by determining the minimum spanning tree associated with the N objects. The minimum spanning tree of the subsupermetric space corresponds to a unique exponential hierarchical tree (hierarchical tree), which can directly determine the supermetric distance matrix $\widehat{d}_{ij}$. Each element in the matrix $\widehat{d}_{ij}$ is equal to the maximum distance between any two adjacent targets when moving along the shortest MST path connecting the starting object to the ending object. Compared with the matrix d_{ij}, there are no more than N - 1 different elements in the supermetric distance matrix $\widehat{d}_{ij}$ [20].

The exponential hierarchical structure tree is based on the working principle of the hierarchical clustering method to cluster the N objects to be clustered and the $N \times N$ distance matrix hierarchically. When studying the relationship between investment strategies of securities market funds, the article believes that funds at the same level have significant common attributes; that is, their actual strategies are the same or similar. The basic steps of the hierarchical clustering method are as follows:

(1) First, the algorithm classifies each object of the system into one category separately and obtains N categories in total. The distance between the categories is the distance between the objects they contain.

(2) The algorithm finds the two categories with the closest distance in the system and merges them into a new category to reduce the number of system categories.

(3) The algorithm recalculates the distance between the new class and all old classes.

(4) The algorithm loops the second and third steps until all the objects in the system are finally merged into one category (this category contains N objects).

The hierarchical clustering method can be divided into single linkage cluster analysis, average linkage cluster analysis, and complete linkage cluster analysis according to the different calculation methods of the distance between clusters in step 3 of the clustering principle. Among them, the distance between clusters in the single linkage cluster method is equal to the minimum distance between two types of objects. The distance between clusters in the average linkage cluster method is equal to the average of the pairwise distances between two types of objects. The distance between classes in the complete linkage cluster method is equal to the maximum distance between the two classes of data. The index hierarchical structure tree in the paper uses the single-link clustering method to reflect the hierarchical relationship of the actual strategies between funds, and the distance obtained by the single-link clustering method is equal to each element of the overmetric distance matrix $\widehat{d}_{ij}$. Figure 1 is the generation process of the minimum spanning tree and the nonunique minimum spanning tree corresponds to the unique exponential hierarchical structure tree.

The Planar Maximally Filtered Graph adds more edges to the minimum spanning tree to ensure that more effective information is retained in the financial network and avoid serious filtering of information on the fund's actual strategy network. The plane maximum filter graph is also based on a complete graph constructed by the distance matrix, and the sum of the distances is made as small as possible under the condition that it is a plane graph. The method of constructing the planar maximum filter graph is similar to the method of minimum spanning tree, with the main differences as follows:

(1) The constraints on the edges are different. MST requires that the edges on the spanning tree cannot appear loops. PMFG relaxes this constraint and requires the edges on the final network graph to be on a plane graph; that is, all edges can be drawn on a plane without crossing out.

(2) The number of edges is different. For a network graph composed of N nodes, there are $N-1$ edges on MST and 3N – 6 edges on PMFG. According to the Kuratowski theorem of plan graphs, PMFG only allows three factions and four factions. The so-called N faction refers to the complete graph containing N nodes in the subgraph of the network graph.

The structure of PMFG is more complex than that of MST; it retains more effective information about the actual fund strategy and is a better supplement to simple MST. Due to the nonuniqueness of the MST method and the seriousness of MST's filtering of fund network information, it is passed. The analysis of topological characteristics of PMFG and MST in time evolution and time scale can verify the effectiveness of the MST method. In addition, the three factions and four factions in PMFG can dig out faction characteristics related to the fund's actual strategy that are not in MST.

The actual fund strategy network represents the relationship between the actual strategies of all funds in the fund market composed of 94 selected funds. From the actual strategy network, the clustering characteristics of funds with the same strategy on the network diagram can reflect the actual fund, whether the strategy is consistent with the declared strategy. When examining the influence of time evolution on the actual strategy of the fund and the change of the time scale on the fund network structure, the position of the fund node in the actual strategy network of the fund will change, and the clustering characteristics of the fund will also be affected. However, this time evolution has an impact on the fund. The degree of influence of the actual strategy and the degree of change of the fund network structure on the time scale are not enough to show from the network diagram. The article calculates the fund network parameters from a quantitative perspective by constructing the financial network topological characteristic index to illustrate the degree of network evolution and the influence of time scale.

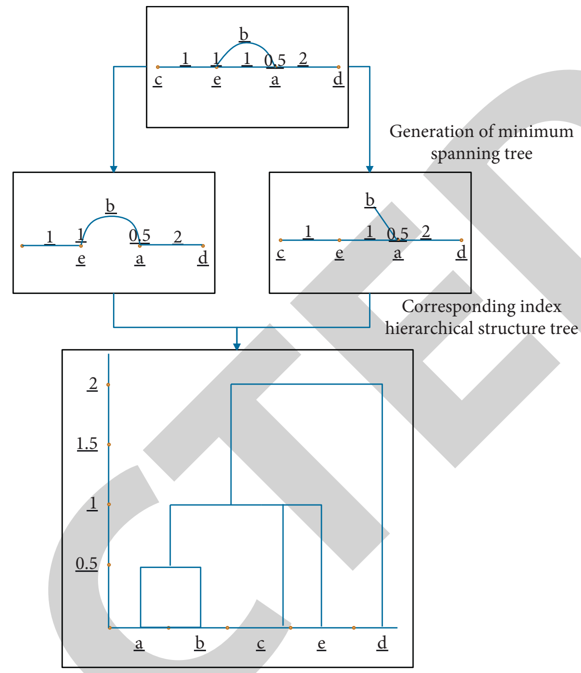


Figure 1: Generation of minimum spanning tree and corresponding index hierarchical structure tree.

The financial network topological characteristic index is the correlation coefficient of the final spanning tree (MST, PMFG), average value, standard tree length, average network path length, average network aggregation coefficient, and central node.

3.3. Analysis of Fund Strategy Network Nodes. The position relationship of nodes on the fund's actual strategy network indicates the degree of correlation between the actual strategies of the funds. The connection of two nodes in the fund's actual strategy network indicates that the actual strategies of the two funds are the same or similar. However, there are many unconnected nodes in the actual strategy network of the fund, and the degree of relevance of their actual strategies cannot be expressed in the actual strategy network of the fund. The article examines the correlation degree of the actual strategy in the network by calculating the mean value ρ of the correlation coefficient between the funds in the actual strategy network of the fund. The larger the mean value of the correlation coefficient ρ, the greater the correlation between the nodes on the fund's actual strategy network, and the higher the actual strategy similarity between funds Equations (11) and (12) are used to calculate the mean values of the correlation coefficients of the minimum spanning tree and the plane maximum filter graph, respectively.

$$\rho_1(\Delta t, T) = \frac{1}{N-1} \sum_{\rho_{ij \in R^{\Delta t}_{e,T}}} \rho_{ij}, \tag{11}$$

$$\rho_2(\Delta t, T) = \frac{1}{3N-6} \sum_{\rho_{ij \in R \Delta t/e,T}} \rho_{ij}. \tag{12}$$

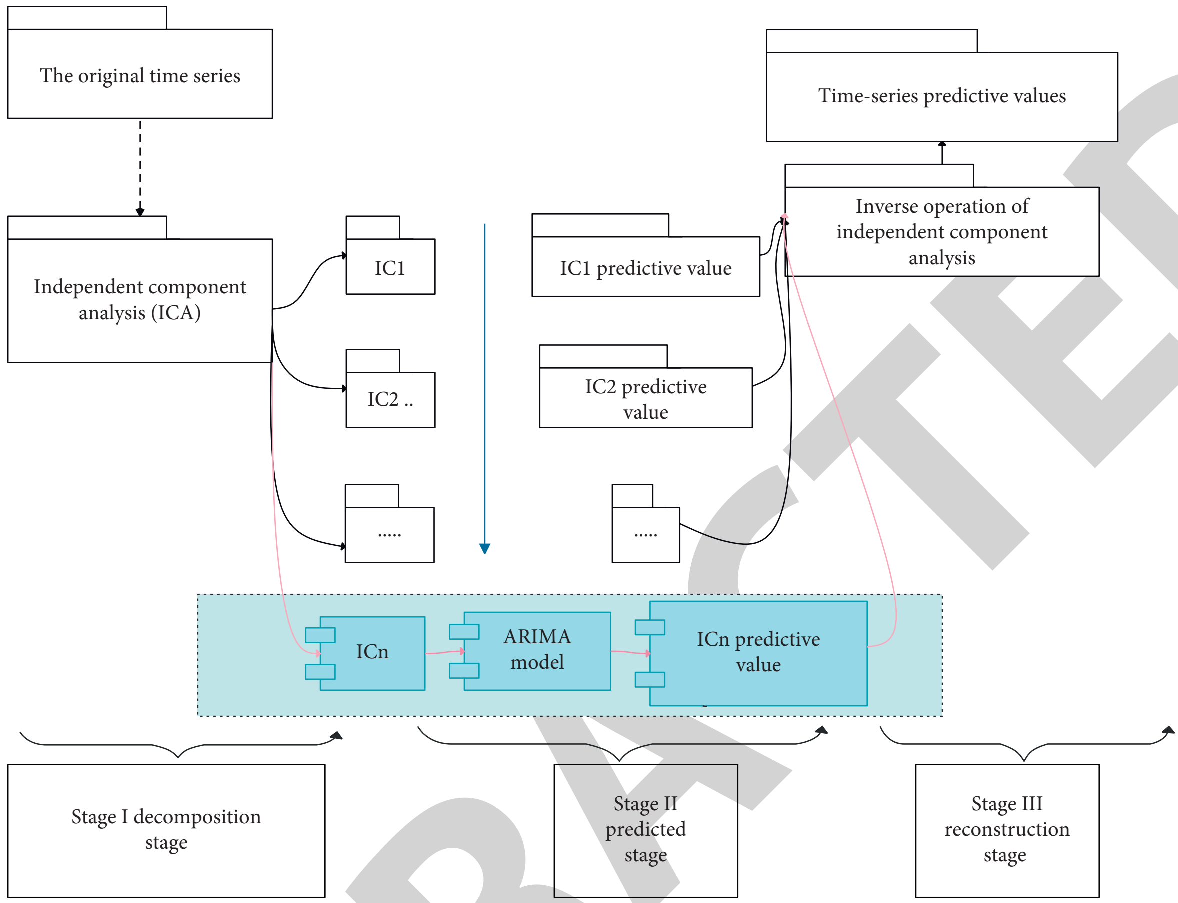


Figure 2: The basic framework of the model.

Among them, $\rho_1(\Delta t, T)$ and $\rho_2(\Delta t, T)$, respectively, represent the mean value of the correlation coefficients of the fund's actual strategy network on the time scale M and the time period T under the minimum spanning tree and planar maximum filter graph methods, and N is the number of fund nodes, and $R_{e,T}^{\Delta t}$ is the correlation coefficient matrix under the condition of the significant correlation coefficient.

In addition, the article uses the correlation coefficient variance $r_1(\Delta t, T)$ and $r_2(\Delta t, T)$ under the minimum spanning tree and the flat maximum filter graph to investigate the actual strategy stability of the fund's actual strategy network. The greater the variance of the correlation coefficient, the smaller the stability of the fund's actual strategy, and the greater the diversity of the actual strategy:

$$r_1(\Delta t, T) = \frac{1}{N-1} \sum_{\rho_{ij \in R_{e,T}^{\Delta t}}} \left[\rho_{ij} - \rho_1(\Delta t, T)\right]^2, \tag{13}$$

$$r_2(\Delta t, T) = \frac{1}{3N-6} \sum_{\rho_{ij \in R_{e,T}^{\Delta t}}} \left[\rho_{ij} - \rho_2(\Delta t, T)\right]^2. \tag{14}$$

The normalized tree length (NTL) is usually used to measure the tightness of the network in financial network theory. The smaller the standard tree length, the tighter the actual strategy network of the fund, and the more similar the actual strategy network between funds. Equations (15) and (16) are the calculation formulas for the minimum spanning tree under the time scale Δt and the time period T and the standard tree length under the flat filter graph:

$$L_{NTL_1}(\Delta t, T) = \frac{1}{N-1} \sum_{d_{ij \in D_{e,T}^{\Delta t}}} d_{ij}, \tag{15}$$

$$L_{\mathrm{NTL}_2}(\Delta t, T) = \frac{1}{3N-6} \sum_{d_{ij \in D \Delta t / e,T}} d_{ij}. \tag{16}$$

Among them, d_{ij} is the Euclidean distance between nodes i and j, and $D_{e,T}^{\Delta t}$ is the correlation coefficient matrix $R_{e,T}^{\Delta t}$ calculated by the distance formula.

The average path length (APL) refers to the average value of the path length between any two nodes. It represents the average number of intermediary edges connecting any two nodes in the network. From the perspective of network theory, it shows that the actual strategies of any two funds in the fund market need to be related at least as many intermediary funds on average.

$$L_{\mathrm{APL}} = \frac{2}{N(N-1)} \sum_{i>j} I_{ij}. \tag{17}$$

Among them, I_{ij}, the path length of any two points, is defined as the number of edges on the shortest path

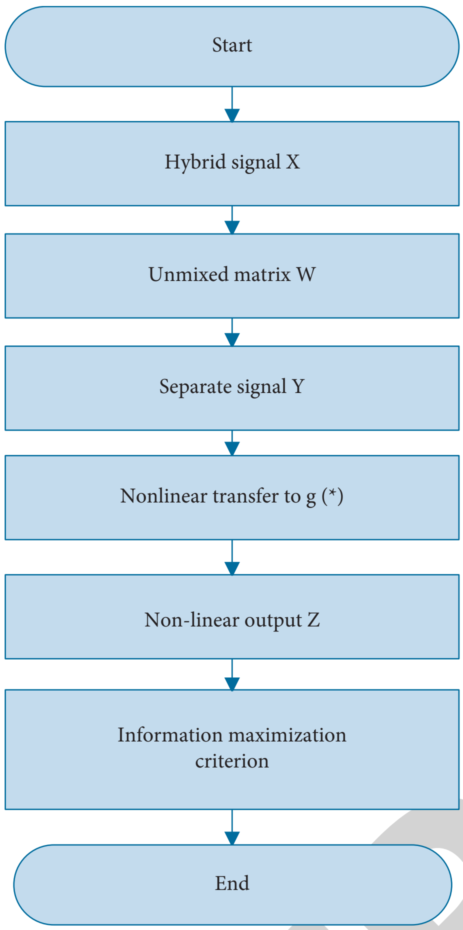


FIGURE 3: Schematic diagram of the algorithm.

connecting two nodes. The smaller the average path length of the fund network is, the more compact the fund network is, and the possibility that the actual strategy similarity between funds will increase increases.

The clustering coefficient C_i of a node is the ratio of the number of edges that are connected to each other in the network of the node set formed by all the nodes connected to the node and the number of edges that constitute a complete graph. The network average clustering coefficient AC is the average value of the clustering coefficients of all nodes on the network.

$$C_i = \frac{2E_i}{k_i(k_i - 1)},$$
$$AC = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{18}$$

Among them, k_i is the degree value of node i, and E_i is the number of edges, where k_i nodes exist in the network.

The degree distribution of each node on the network graph is not uniform. Some nodes have a large degree value, while other nodes have a small degree value. This feature is obvious on the interpersonal network. People usually rank the node with the highest degree value, defined as a key node.

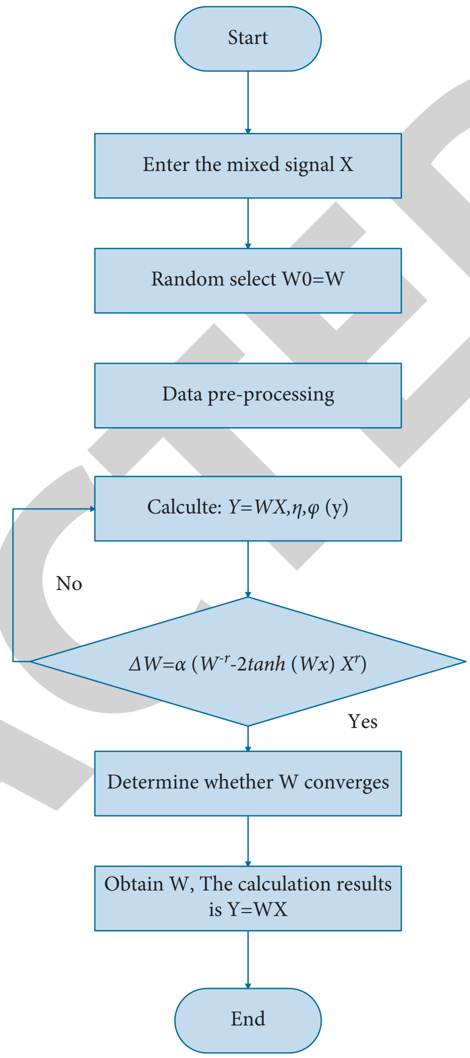


FIGURE 4: Flowchart of the algorithm.

Because the degree of key nodes is very large and the correlation with other nodes is great, key nodes can often reflect the characteristics of the network. In the article, the nodes with the highest node degree in the network are considered key nodes.

The actual strategy of nodes on the fund's actual strategy network is constantly changing over time. Fund managers will change their actual strategies according to the market situation and their own technical means. Regarding the fund market as a network, there are correlations between nodes. Key nodes represent nodes with a greater degree of correlation between the fund and the actual strategies of other funds, representing the most common actual strategies in the fund market, and it can reflect the overall fund market and the change trend of the actual strategy over time. It is worth noting that it is not the biggest influencer of the actual strategy of the fund's actual strategy network. It cannot affect the changes of the actual strategy of other funds with the same or similar strategies but the actual strategy of the actual strategy network and the best representative of time evolution.

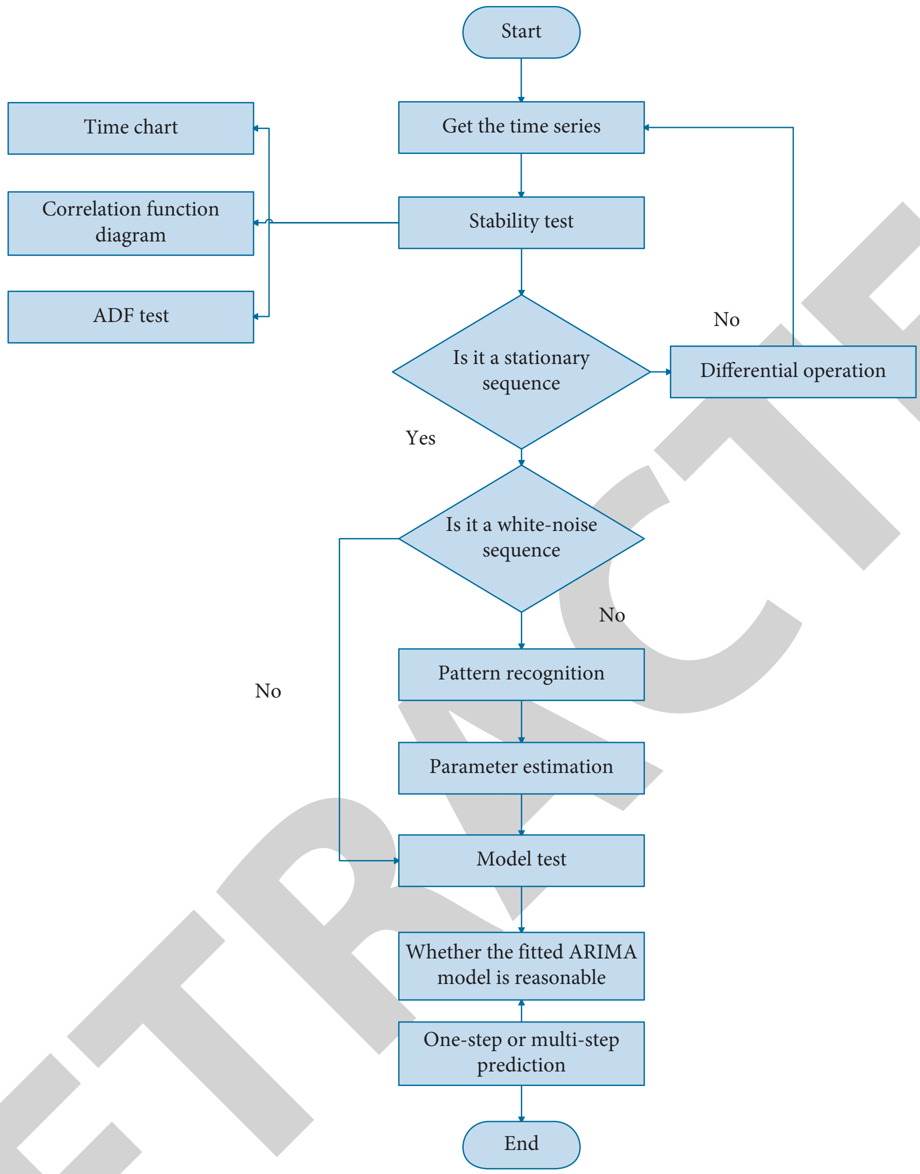


FIGURE 5: The modeling flowchart of the ARIMA model for each independent component.

4. Analysis Model of Influencing Factors of Economic Fluctuation Based on Big Data

4.1. System Model Building. The model is mainly divided into three stages in the forecasting process. The first stage is the decomposition stage. Its main task is to find out the potential factors that cause time series fluctuations, to effectively separate independent source signals through independent component analysis algorithms, and to analyze economic meanings simply in combination with their fluctuation patterns. The second stage is the prediction stage. Its main task is to find the optimal ARIMA model to predict each independent component based on its own data characteristics. The third stage is the reconstruction stage. Its main task is to reconstruct the predicted value of each independent component through a certain mechanism to form the predicted value of the original time series, and this mechanism is the inverse operation of independent component analysis. The framework of the model is shown in Figure 2.

Figure 3 describes the principle of the Infomax algorithm and draws a flowchart. The core idea of the algorithm is to measure the independence between various variables through information entropy and apply it to the time series observation data by selecting a nonlinear function so as to realize the separation of the independent components of the time series.

The calculation process of the Informax algorithm is drawn as the following flowchart in Figure 4.

The modeling flowchart of each independent component ARIMA prediction model in the prediction stage of the ICA-ARIMA model proposed in this article is shown in Figure 5.

In order to avoid the impact of different sample data sets on the experimental results, the experiment will be repeated

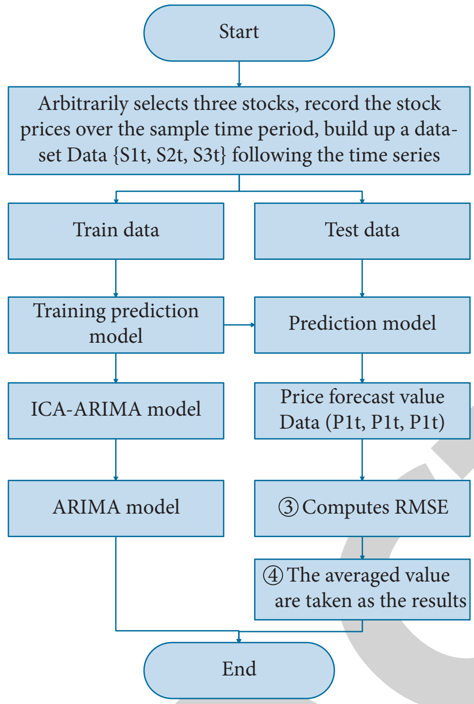


Figure 6: The comparative experimental process of the ICA-ARIMA model and ARIMA model.

Table 1: Statistical table of economic volatility index.

Index	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20
Comprehensive PMI index	53	28.9	53	53.4	53.4	54.2	54.1	54.5	55.1	55.3
Manufacturing PMI	50	35.7	52	50.8	50.6	50.9	51.1	51	51.5	51.4
Nonmanufacturing PMI	54.1	29.6	52.3	53.2	53.6	54.4	54.2	55.2	55.9	56.2
Index	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21
Comprehensive PMI index	55.7	55.1	52.8	51.6	55.3	53.8	54.2	52.9	52.4	48.9
Manufacturing PMI	52.1	51.9	51.3	50.6	51.9	51.1	51.0	50.9	50.4	50.1
Nonmanufacturing PMI	56.4	55.7	52.4	51.4	56.3	54.9	55.2	53.5	53.3	47.5

until all the stocks in the bank stock pool are selected at least once. The final result will also be displayed by taking the average of the times of all data sets. The specific experimental process is shown in Figure 6.

4.2. Analysis and Discussion. In order to verify the effectiveness of the model proposed in this article in the economic market, this article selects the PMI index to form the relevant experimental data set to carry out the experiment and conducts the simulation test in Matlab.

On the basis of the above analysis, this article combines the PMI index to carry out the application of big data mining technology in the analysis of the influencing factors of economic fluctuations, calculates the PMI index in recent years, and draws it into a statistical chart, as shown in Table 1 and Figure 7:

On the basis of the above analysis, the analysis results of economic fluctuation factors are counted, and the prediction accuracy of the factors affecting economic fluctuations is calculated. The system in this

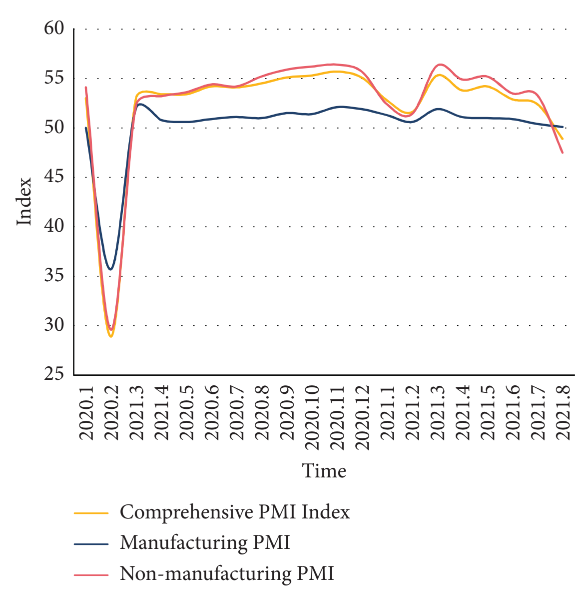


Figure 7: Statistical diagram of economic fluctuations.

Table 2: Verification of the effect of data mining in the analysis of the influence factor of economic fluctuations.

Number	The method of this paper	The method of [17]	Number	The method of this article	The method of [17]
1	89.66	87.05	31	83.69	80.08
2	92.48	83.61	32	83.47	77.48
3	89.71	85.22	33	88.03	83.14
4	86.64	84.31	34	82.91	80.02
5	80.98	77.03	35	83.94	83.78
6	80.25	75.88	36	91.07	88.90
7	84.30	77.45	37	87.84	82.05
8	82.51	74.94	38	87.35	82.25
9	84.84	79.35	39	84.07	80.11
10	83.29	76.37	40	85.95	79.29
11	87.96	80.05	41	81.94	75.63
12	92.42	91.63	42	88.51	84.17
13	93.24	89.00	43	88.55	84.83
14	81.27	73.97	44	84.15	82.70
15	92.22	92.03	45	83.44	80.24
16	80.24	74.49	46	92.20	83.02
17	91.11	89.42	47	84.51	77.75
18	90.80	89.33	48	80.66	76.64
19	92.55	87.86	49	88.07	84.80
20	93.68	91.64	50	82.62	77.61
21	84.39	82.12	51	85.56	85.47
22	87.79	80.46	52	87.66	87.39
23	81.03	77.52	53	80.86	75.74
24	85.17	77.67	54	92.48	89.79
25	83.65	75.31	55	83.99	83.12
26	92.45	83.90	56	86.62	81.19
27	84.03	78.63	57	88.86	80.87
28	88.75	88.07	58	92.79	85.57
29	82.93	81.74	59	85.21	79.71
30	80.82	72.97	60	92.99	87.30

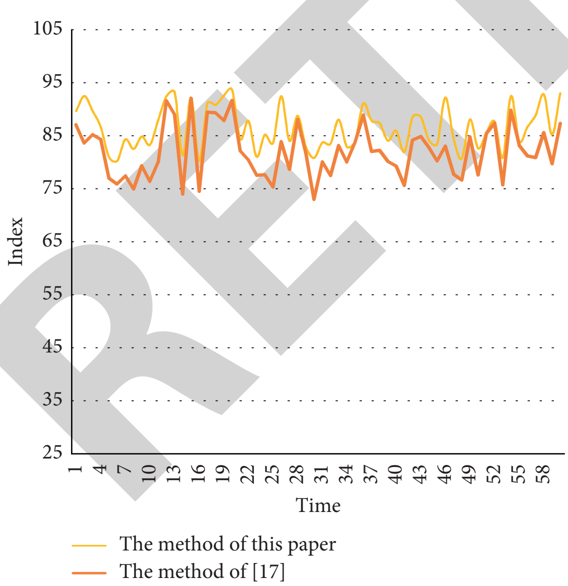


Figure 8: Analysis of the influencing factors of economic fluctuations in the era of big data.

article is compared with the literature [17] through the simulation platform, and the statistical effect of the statistical economic fluctuation index and the analysis effect of the influencing factors of economic fluctuation are compared, and the results are shown in Table 2 and Figure 8.

It can be seen from the above research that the data mining iteration proposed in this article can play an important role in the analysis of the influencing factors of economic fluctuations and accurately dig out the relevant factors that affect the economy.

5. Conclusion

Data mining often selects historical data for the statistics of keyword frequency, and it is difficult to estimate the economic uncertainty of the current or even forecast the future. According to the efficient market hypothesis, all kinds of information will be absorbed by all markets at the same time in the shortest time, so there is no spillover effect between financial markets. In fact, with the deepening of research, there is a widespread spillover effect in financial markets. According to existing research, the transmission of information through yield or volatility channels is called mean spillover and volatility spillover, respectively. Mean spillover refers to changes in asset prices or returns in one market that affect changes in asset prices or returns in other markets. Volatility spillover refers to the volatility of asset prices in one market affecting other market volatility, which is generally measured by the conditional variance of prices or returns. This article proposes an analysis model of the

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9821647, 1 page
https://doi.org/10.1155/2023/9821647

Retraction

Retracted: The Impact of Financial Development on Agricultural Enterprises in Central China Based on Vector Autoregressive Model

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Kang and M. Zhao, "The Impact of Financial Development on Agricultural Enterprises in Central China Based on Vector Autoregressive Model," *Security and Communication Networks*, vol. 2022, Article ID 5629202, 16 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 5629202, 16 pages
https://doi.org/10.1155/2022/5629202

Research Article

The Impact of Financial Development on Agricultural Enterprises in Central China Based on Vector Autoregressive Model

Jian Kang[1,2] and Minjuan Zhao[1]

[1]College of Economics and Management, Northwest A&F University, Yangling, Shaanxi 712100, China
[2]Center for Shaannan Eco-Economy Research, Ankang University, Ankang, Shaanxi 725000, China

Correspondence should be addressed to Minjuan Zhao; minjuan.zhao@nwsuaf.edu.cn

Received 18 January 2022; Revised 23 February 2022; Accepted 7 March 2022; Published 31 March 2022

Academic Editor: Chin-Ling Chen

China is a large developing country, and there is no doubt that it is a large developing agricultural country. Agriculture has played a vital role in the development of the national economy since ancient times, and it serves as the foundation of China's national economy. As agricultural modernization is the fundamental way of agricultural development, its source of funds must come from rural finance. The financial support of rural finance for agricultural modernization is the material basis and prerequisite for ensuring the healthy and rapid development of agricultural modernization. If there is no financial support from rural finance, the process of agricultural modernization will be hindered, and it will not benefit our country's socialist modernization. Although our country's rural financial system has been constantly changing and improving, our country's current financial system still has many shortcomings and defects. In the process of supporting agriculture, rural finance has insufficient capital supply, which has become a "bottleneck" in the development of agricultural modernization, and to a certain extent has severely restricted the development of agricultural modernization. On the basis of existing literature research, this article systematically studies the influence of rural financial market structure on the development of agricultural enterprises in central China. The content includes agricultural total factor productivity and agricultural emission reduction. When analyzing the impact of rural financial market structure on agricultural development, this article also pays attention to its specific impact on agricultural development under the conditions of different types of enterprises.

1. Introduction

Every year, China issues documents stipulating preferential policies to support the development of leading agricultural enterprises, and the Agricultural Development Bank formulates corresponding loan measures for leading agricultural industrialization enterprises in accordance with relevant national policy documents. The national policy has initially taken specific measures to support the loans of leading enterprises and then transformed them into relevant regulations. The Agricultural Development Bank is formulated in accordance with relevant national policy documents. Correspondingly, it provides loan methods for leading enterprises in agricultural industrialization. In addition, the scope and intensity of support have been expanded this week, reflecting the state's emphasis on and strong support for the development of leading enterprises. The national policy has changed from the initial specific measures to support leading enterprise loans to the relevant provisions to expand the scope and intensity of support, which reflects the state's emphasis on and strong support for the development of leading enterprises [1–3]. The Agricultural Development Bank's loan methods for leading enterprises are gradually supplemented and improved, and the loan amount has been increasing year by year. With the support of China's policies and the strong support of the Agricultural Development Bank, most of our country's leading agricultural enterprises have developed rapidly.

There are more and more leading agricultural enterprises, and the economic benefits created are increasing. As the world's largest agricultural country, the development of China's rural financial market structure has evolved from a complete monopoly of the People's Bank of China to a monopolistic competition that tends to be more competitive [4–7]. The Agricultural Development Bank has gradually supplemented and improved the loan methods for leading enterprises, and the loan amount has increased year by year. Every year, China issues documents stipulating preferential policies to support the development of leading agricultural enterprises. With the support of national policies and the strong support of the Agricultural Bank of our country, most of China's leading agricultural enterprises have developed rapidly. With the deepening of the reform of the country's financial system, China's rural financial market is changing in terms of market structure types, rural financial market's main asset investment scale, operating performance, regional network layout, and the impact of rural financial market structure on agricultural development. And there are strong regional development differences [8–10].

The impact of the development and adjustment of the rural financial market structure on agricultural development is closely related to the resource endowment structure and development level of agricultural production factors in rural areas in China. These agricultural production factors' resources are related to the country or region's openness, agricultural development policies, and agricultural functions [11–13]. The regional and regional development environment and other influencing factors jointly determine the type and development characteristics of China's agricultural industrial structure. Therefore, how to find a matching rural financial market structure under the constraints of the existing agricultural economic structure and the harmonious development of the two will provide a stronger internal economic driving force for the structure of the rural financial market to positively influence the sustainable and efficient development of China's agriculture [14–17]. Bank loans are mainly carried out and formulated in accordance with relevant national policy documents. As a key work content, agricultural industrialization has been widely promoted in different regions. In order to further expand the scope and degree of support, the state focuses on the corresponding assistance work for the development leading enterprises identified by the state. The national policy has gone through a series of processes from the proposal, to the promotion of the policy, to the implementation of the policy. Although the proportion of its gross output value in the gross national product has been declining in recent years, its position in the entire national economy is still very important. The development of agriculture has made significant contributions to the development of the entire national economy and social progress in our country. Agriculture occupies an important position in the entire national economy of our country. However, due to our country's policy of giving priority to the development of industry in the early days of the founding of the People's Republic of our country, our country's agricultural development was relatively lagging, and the development of agricultural modernization was naturally relatively backward. So far, agricultural modernization, as a key link in the modernization drive, seems to have become a shortcoming in our country's modernization drive [18–21].

As the world's largest agricultural country, China's rural finance has developed from the initial monopoly of a financial institution to the current diversified development of multiple rural financial institutions. The development of the rural financial market structure has experienced a trend from a complete monopoly of the People's Bank of China. The type of monopolistic competition with strong competition evolves. With the deepening of the reform of the country's financial system, China's rural financial market is changing in terms of market structure types, rural financial market's main asset investment scale, operating performance, regional network layout, and the impact of rural financial market structure on agricultural development. And there are strong regional development differences [22–24]. There are more and more leading agricultural enterprises, creating higher and higher economic benefits. As the world's largest agricultural country, China's rural financial market has rapidly expanded and evolved from the complete monopoly of the People's Bank of China to competition among the more intensely competitive subunits. With the deepening of the national financial system reform, our country's rural financial market has changed in the type of market structure. The impact of the development and adjustment of the rural financial market structure on agricultural development is closely related to the resource endowment institutions and development levels of agricultural production factors in rural areas in China. Development policies, agricultural functional areas, regional development environment, and other factors jointly determine the types and development characteristics of China's agricultural structure. Therefore, how to find a matching rural financial market structure under the constraints of the existing agricultural economic structure? The harmonious development of the two will provide the internal economic driving force for the rural financial market structure to influence the sustainable and efficient development of China's agriculture.

From the triumphant advancement of China's agricultural development since the reform and opening up to the current sound development trend of steady progress, it has provided a solid foundation for the steady development of the country's economy and society. The agricultural structure has been further optimized, the agricultural development mode has continued to change, and the agricultural development efficiency has been steadily improved. However, China's agricultural development is currently facing some severe challenges, such as the slowdown of the overall economic growth, the imbalance of the supply and demand structure of agricultural products and the low quality of agricultural products, the difficulty of increasing the income of the main agricultural operators, and serious agricultural pollution. As shown in the 2017 National Economic and Social Development Statistical Communiqué, the added value of the primary, secondary, and tertiary industries increased by 3.9%, 6.1%, and 8.0%, respectively; the primary, secondary, and tertiary industries accounted for the respective proportions of GDP They were 7.9%, 40.5, and

51.6%. Therefore, regardless of the agricultural growth rate and the contribution to the national economy, there has been a serious weakening. On the one hand, it is related to the inherent weakness of agricultural development and the impact of more natural resources, climate, and other uncertain factors. On the other hand, this is also related to the long-term focus on high-input and extensive development of agricultural production factors [25–28]. High consumption, high pollutant discharge, low efficiency, and low output are directly related. Therefore, based on the existing basic conditions and factual characteristics of the rural financial market structure and agricultural development, it will be more beneficial to analyze the key factors or focus of rural finance that affect agricultural development and to combine the regional differences between China's rural financial market and agricultural development. The article makes a detailed analysis of the research block diagram. The research framework of this paper contains two aspects. On the one hand, policies related to agricultural development mainly involve external environmental factors such as natural resources and climate resources; on the other hand, the withdrawal of agricultural-related bank loan policies is also related to the long-term production factors. High investment is related to the development of extensive expansion forms. The main influencing factors involved include high pollution, high energy consumption, low output, and low efficiency. The optimization and adjustment of rural financial market structure and agricultural industrial structure, and it is of great significance to explore the best force point for the rural financial market structure to affect agricultural development [29]. The research logical structure of this paper is shown in Figure 1.

2. Existing Research Results and Theoretical Basis

2.1. Financial Deepening and Financial Repression Theory. Research on the relationship between financial development and economic growth. Some scholars pointed out that the sound development of a country or region's financial sector is an important factor in promoting the country's economic development. Some scholars also believe that in studying the correlation between finance and economic growth, there are two analytical directions: "demand following" and "supply leading". The former emphasizes the service demand of economic entities on the financial market, and the development of the financial market is the inevitable result of economic growth, while the latter emphasizes the issue of supply priority related financial service support, and the support of financial services promotes economic growth. These two different development models are also closely related to the economic structure at different growth stages. In the economic society in the early stage of development, the supply-leading financial industry is in a leading position in the market, and the demand-following financial model is at the core when the economy grows to a high-level development stage. Therefore, the two development models of "demand following" or "supply leading" indicate that the interactive influence of financial and economic development has not formed a virtuous circle. Based on the macroeconomic data of 35 countries, some scholars have concluded that "there is a roughly parallel development link between economic growth and financial market development", but this conclusion does not have a relationship between various factors [30–32].

The core viewpoints of financial deepening and financial repression theory are as follows.

(1) There is a positive interaction between the structure of the financial market and economic development. Due to the lack of institutions and the inadequacy of government supporting policies, the development of underdeveloped countries and regions may be completely another result: the underdeveloped countries have excessive market intervention in all aspects of economic and social activities; excessive financial systems and regulations affect the sustained and healthy development of the financial market structure. Research on the relationship between financial development and economic growth exists. Existing research shows that the healthy development of the financial industry is an important factor in promoting the economic development of a country or region. In addition, some scholars believe that, at the current stage, we are conducting research on the correlation between finance and economic growth, including two main details and directions of "demand following" and "supply leading". The lagging development of the financial industry will inhibit economic growth, and the lagging economic growth will hinder the sustained development of the financial market, which in turn creates a vicious circle of lagging financial market development and economic development difficulties [33].

(2) To give full play to the positive role of the financial market structure in promoting economic development, the problem of "financial repression" must be resolved and the process of "financial deepening" must be promoted. Galba's and other scholars supplemented and perfected the core content of McKinnon and Shaw's financial repression and financial deepening from different research perspectives. On the one hand, the demand for services emphasizes the service demand of economic entities for the financial market. The entire operation route framework mainly includes the following three aspects. The focus is on the supply and demand of various economic entities to the financial market. First of all, the study believes that the initial development of the financial market is the primary prerequisite for economic growth, and economic growth emphasizes the priority supply support of financial services. The development of financial services has also indirectly contributed to the growth of the entire economic entity. In the technical scheme, three different modes of economic expansion will also be linked with the stage of economic development and the structure of economic

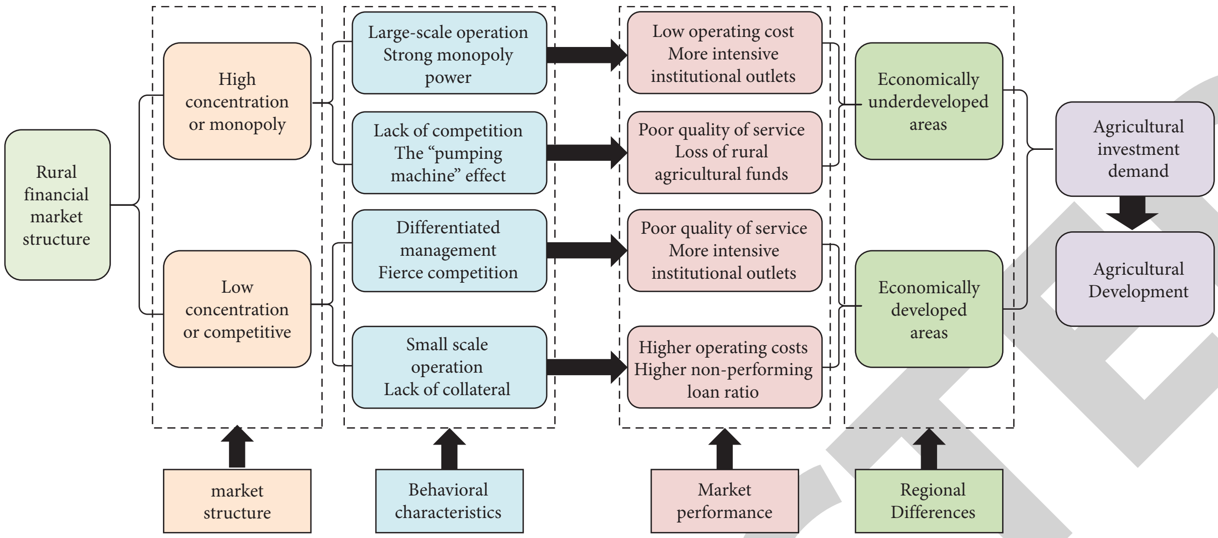


Figure 1: The research logical structure of this paper.

development. In the initial stage, the supply-led financial industry has always been in a leading position, which in turn provides support for relevant supporting policies in the financial sector. The development of financial markets is an important part of economic growth. On the other hand, supply priority and financial service support are also issues that we need to focus on. The support of financial services promotes economic growth. They pointed out that the implementation of low market interest rates will lead to the excessive pursuit of funds such as rural credit by fund demanders, which will lead to financial markets. Nonprice rationing causes double price distortions in the financial market and economic market; raising market interest rates can reduce the unnecessary demand of nonefficient investors for agriculture-related funds and increase the marginal rate of return on investment demand. Model analysis results are shown in Figure 2.

2.2. Agricultural Finance Constraint Theory. In the 1990s, some countries such as East Asia did not adopt financial market liberalization policies but achieved significant economic development achievements, while countries such as Latin America did the opposite. However, the fragile financial market system made economic development difficult. Even if the "financial deepening" theory is perfected and revised, there are still many flaws. Unreasonable financial liberalization measures are excessively taken against developing countries, regions, and transition economies. As pointed out by the "financial deepening" theory, the government states that product pricing and quantity control in the financial market are the direct causes of distortions in the allocation of resources in the financial market, which is exactly the opposite of the successful experience of some East Asian countries. Regardless of East Asian developed countries such as Japan and South Korea, or underdeveloped countries such as China and Malaysia, although there are varying degrees of financial repression, they have all achieved rapid economic development. On the contrary, due to countries that have adopted excessive financial liberalization, such as Chile and Argentina, the results are exactly the opposite of the expected results [34].

The mechanism of financial restraint theory is as follows. First, the prerequisites for the realization of financial restraint theory are that a country or region has an overall stable economic policy environment, low price levels, and positive real market interest rates for deposits and loans. The important difference between financial restraint and financial repression is that the government does not grab rents from the private sector but creates rents for the private sector. Second, financial constraints complete rent transfer through government rent creation and competitive activities of economic entities. Rent is the link between the government and the private sector. The government in the theory of financial restraint only creates opportunities for private sectors to obtain rents. To obtain rents, competition among economic actors is required. Third, in the theory of financial constraints, the government ensures that financial market operators obtain stable returns through institutional arrangements such as reasonably floating interest rate interventions on deposits and loans, setting entry barriers, conditional asset substitution, and creating a predictable investment and financing environment. Inhibition is that government departments harvest rents so that the nominal interest rate is below the inflation rate and maintains a negative real interest rate. The economic structure of different growth stages is also closely related to a variety of different development models. In the early stage of development, the economy and society need more help from supply-led financial enterprises, which will help them enter a market-leading position. Taking into account the fact that the financial model adapted to the enterprise is the core competitiveness of the economic growth to a high level of

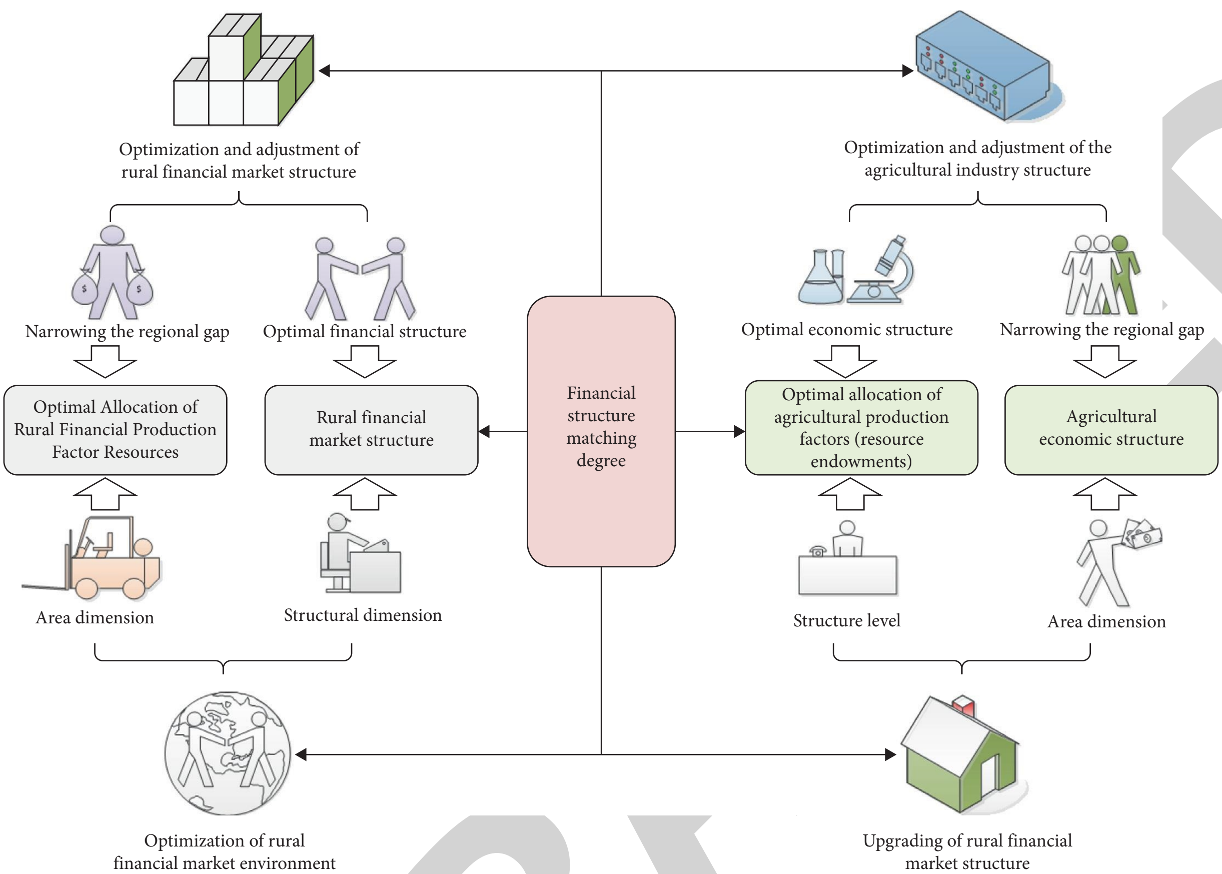


Figure 2: Rural financial market influences the mechanism of agricultural development.

development, therefore, we comb through the two development models of "demand follow" or "supply lead" and find that the interactive impact of finance and economic development has not yet formed a virtuous circle. The purpose of government departments' financial constraints is to create more rental opportunities and incentives for the financial sector and capital demanders. Banks and other financial institutions provide sufficient credit funds for the "agriculture, rural areas, and farmers" capital demanders while obtaining more development opportunities, thereby reducing a series of uncertain issues caused by information asymmetry [35].

The breakthroughs of the theory of financial constraints are as follows. First, it is believed that the government's moderate intervention and market regulation can be coordinated. The government must create rental opportunities for the various operating entities of the financial system through various effective institutional arrangements to encourage the efficient operation of the entire financial system. Second, the theory is based on the particularity of financial deepening in developing countries, while intervening in interest rates in an orderly manner. At the same time, some scholars have come to the following main conclusions by analyzing the macroeconomic data of 35 countries; "there is a roughly parallel developmental relationship between financial market development and economic growth". However, these conclusions show that there is no relatively significant comparative relationship among the factors. In addition, both financial deepening and financial repression theories have corresponding core views, which are reflected in the positive interaction between financial market structure and economic development. To open up the financial market locally is not to blindly promote interest rate liberalization. Third, compared with the theory of financial repression, it is clear that the government's moderate intervention is to create rents for the private sector, not to extract rents. The interrelationships between model elements are shown in Figure 3.

2.3. Rural Financial Development Theory. With the continuous deepening of research on rural finance-related issues, three representative theories have emerged in the rural financial development theory, namely, the theory of rural financial regulation, the theory of the rural financial market, and the theory of imperfect competition in rural finance. Suppose there are agricultural operators, especially poor farmers who do not have the ability to save, and the supply of agricultural development funds is insufficient. Due to the low profitability of agriculture, the long-term nature of the demand for agricultural credit funds, and the uncertainty of agricultural income, it indicates that agriculture cannot become the financing object of commercial banks. However, due to the lack of institutions and government support

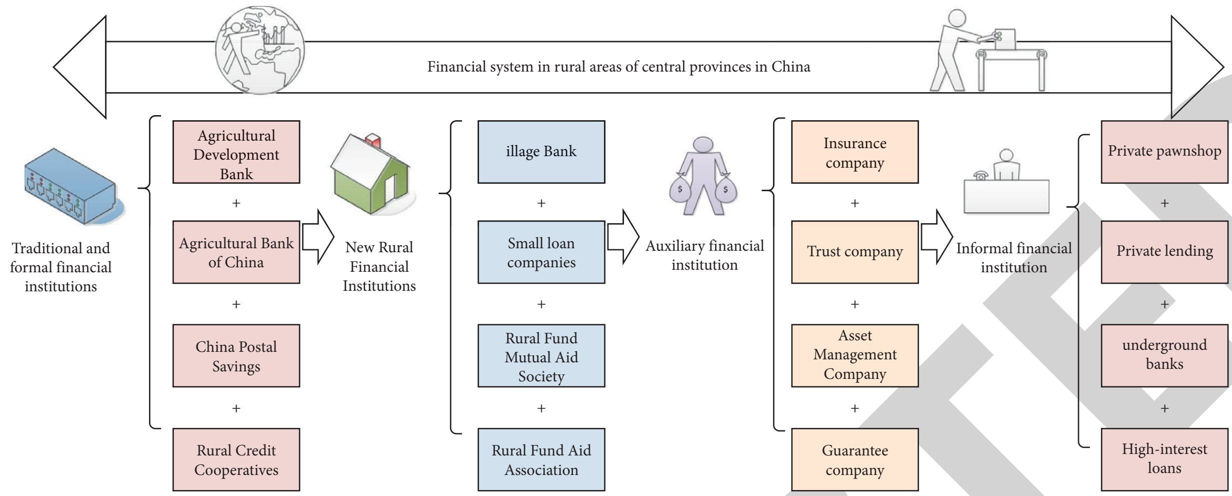


Figure 3: The influence model of rural financial market structure on agricultural growth.

policies, the development of less developed countries and regions may be completely different results. There are excessive market interventions in all aspects of economic and social activities in less developed countries. Excessive financial systems and regulations affect economic development. This ensures the sustainable and healthy development of China's financial market structure. The lagging development of the financial industry will inhibit economic growth. Therefore, commercial financial institutions with the goal of maximizing profits have insufficient motivation to enter the rural financial market, and they advocate government intervention in the rural financial market and occupy a leading position in the financial market. It is concluded that the government should carry out necessary interest rate control and market access control, and the state should set up special nonprofit rural financial institutions to inject policy-based low-interest-rate funds into rural financial institutions to increase agricultural production input and alleviate rural poverty. In order to support the development of the rural agricultural economy and strictly restrict informal finance in rural areas. A large number of low-interest policy funds should be injected into rural agriculture through rural branches of banks and agricultural credit cooperative organizations.

During the development of a country's financial system, with the development of the economy, the financial-related ratios tend to increase; the financial-related ratios of economically underdeveloped countries are usually much lower than those of developed countries; the development of financial markets starts with traditional banks. The industry continues to mature and depends on the transmission and circulation functions of paper money in the economy and society; with the development of economy and society, banks, as part of the financial market, will tend to reduce the proportion of the total assets of a country's financial institutions. And the proportion of other new financial institutions will rise accordingly; for the financial development of most countries, the exemplary role of advanced countries is as important as international capital flows.

According to the cost-benefit criterion, choose the financial market and demand realization mechanism that can meet the economic system's demand for financial service functions as a platform to realize the functional coupling between the financial system and the external environment. The financial constraint model involves the following aspects. First of all, the theoretical realization of financial constraints is based on the long-term stable economic policy environment of countries and regions. Therefore, compared with other influencing mechanisms, generally lower price levels and positive market storage rates have become the main influencing factors. In addition, the rent factor caused by the country's fiscal restraint policy and the correlation factor between the government and the real economy have also become important influencing factors. The competition among different development entities shall be intervened in deposits and loans through reasonable floating interest rates, setting entry thresholds, conditional asset substitution, and creating a predictable investment and financing environment. It has also become one of the structures of the financial model. The core ideas of the evolution of financial market structure are as follows. First, the financial service function of financial institutions has become more stable, and financial functions have changed relatively little in different periods and different political regions; second, a more efficient financial system is the inevitable innovation of financial institutions and market competition result. In practice, many Asian countries have shown that poor farmers also have savings needs. If savings incentives are implemented, the poor will increase their savings. The way to reduce poverty is to establish an effective and sustainable rural financial market operation mechanism, not to increase loans or savings. If the agricultural credit subsidy policy is implemented, it will be difficult to build an efficient and sustainable rural financial system. Agricultural credit subsidies are only applicable to areas where the rural financial market mechanism is seriously malfunctioning.

3. The Influence of the Financial Market on the Structure of Rural Enterprises

3.1. The Influence of Financial Market Structure on the Growth of Agricultural Enterprise Output Value. Existing rural financial market theories have their own advantages and shortcomings in analyzing the impact of rural financial market structure on agricultural development in different countries or regions. As the economy develops, financial correlation ratios tend to rise, and a country's financial system also continues to evolve. In addition, financial correlation rates in less developed countries are generally much lower than in developed countries. The development of the financial market begins with the continuous maturity of the traditional banking industry, which relies on the dissemination and circulation of paper money in the economy and society. These theories do not matter about the suppression and deepening of rural finance, the dominant mode of supply and demand in the rural financial market, and the relationship between rural financial structure and rural financial structure. The financial function is still based on the theory of rural credit and modern rural financial market, most of which are based on the development of rural financial system institutions or rural financial market structure itself but rarely consider the development of a country or region's agricultural economic structure and regional spatial structure.

Without considering noneconomic structural factors such as the natural geographic attributes of a country or region and regional spatial differences, but in terms of the characteristics of the development of the rural financial market structure, the advantages and disadvantages of different types of rural financial market structures in the process of agricultural development coexist. In a country or region, if there is a rural financial market structure with a relatively high degree of market concentration (strong monopoly), it will have a positive impact on agricultural development.

Linear regression mainly reflects the linear relationship between predictive variables and independent variables, which can be expressed as

$$E(Y) = a + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n. \tag{1}$$

Extra relative risk (ER) is used to reflect the effect of environmental factors on sports health risks. The ER value is based on the relationship coefficient in the regression model β. The calculation formula is

$$\begin{aligned} \mathrm{RR} &= \exp(\beta x), \\ \mathrm{ER} &= (\mathrm{RR} - 1) * 100, \\ \mathrm{ER}(95\%\mathrm{CI}) &= [\exp[(\beta \pm 1.96se)x] - 1] * 100. \end{aligned} \tag{2}$$

Among them, B_i is used as the scale element corresponding to the i-th evaluation in data set B. Through data set B, the data membership vector representing the injury of the athlete can be effectively integrated into a scalar. The formula is expressed as

$$V = r_j \times B. \tag{3}$$

It is mainly reflected in the following. First, rural financial institutions themselves can rely on economies of scale to carry out business (such as bank credit) or through moderate scale expansion can effectively reduce operating costs such as transaction costs and expenses and improve their own operating efficiency. Market performance thereby effectively allocates credit funds. Second, rural financial institutions (such as China's large state-owned commercial banks and rural credit cooperatives) with high monopoly market power are mobilizing savings and identifying borrowers—the credit rating of agricultural business entities—and to diversify risks, it will be easier to acquire high-end and high-quality borrowers as potential customers and increase credit supply to high-quality borrowers in the market. Third, rural financial institutions that are in a market monopoly position are more likely to take advantage of their financial market power. Not only can it rely on the advantages of scale to convey the higher credibility of the company to potential depositors in the market, but it is also easier to reach long-term cooperation agreements with borrowers such as large agricultural operation organizations, leading agricultural industrialization enterprises, or agricultural operation entities with higher credit. With the development of the economy and society, banks are gradually integrated into the existing national financial market. The proportion of banks will tend to reduce the proportion of the total assets of a country's financial institutions. However, the proportion of other emerging financial institutions will also increase accordingly. For most countries, the development of the financial industry also reflects the complementarity of financial capital and active development in developed countries. It is conducive to increasing the loan line of agricultural borrowers and then can give full play to the basic financial functions of rural banking and other financial institutions such as mobilizing savings, allocating credit funds, and controlling risks. Model analysis results are shown in Figure 4.

At the same time, the relatively high degree of market concentration (strong monopoly) of the rural financial market structure has adverse effects on agricultural development mainly reflected in the following: Banks (rural credit cooperatives) due to the lack of effective competition and lack of motivation to innovate financial products improve their own service quality and improve business models. According to the cost-benefit criterion, we generally choose financial service functions that can satisfy the economic system. The needs of the financial market are mainly concentrated in the construction of the realization mechanism platform, so as to realize the functional coupling between the financial system and the external environment. Second, in the rural financial structure with a strong monopoly, rural financial institutions are more aggressive. Earnings or avoiding business risks, market behaviors such as lowering service quality, setting higher prices of financial products, and raising credit thresholds are often adopted, resulting in the inability to effectively allocate credit resources to more

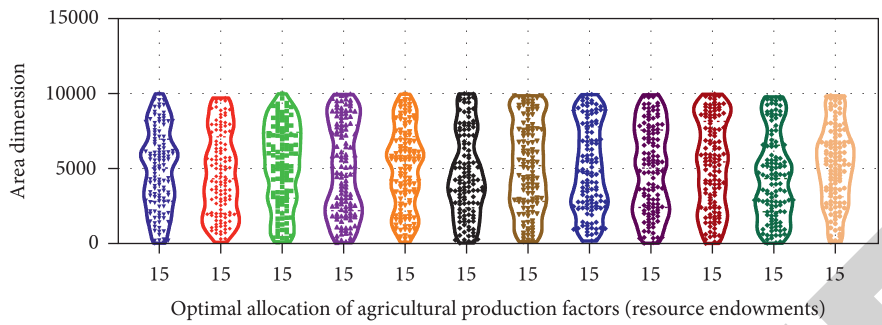


FIGURE 4: Statistics on the distribution of leading agricultural enterprises in various industries.

agricultural business entities. Large commercial banks are operating in rural agriculture. The market behavior of "credit gracious" once became the rural financial market.

Shape the general functional relationship between the output y of the injury model and the input $x_1, x_2, \ldots, x_n$. The Kolmogorov-Gabor polynomial is as follows:

$$y = f(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2. \tag{4}$$

And treat each of the monomials as m input models in the original structure of the modeling network:

$$\begin{aligned} v_1 &= a_0, \\ v_2 &= a_1x_1, \\ v_3 &= a_2x_2, \ldots, v_6 = a_5x_1x_2. \end{aligned} \tag{5}$$

The final information $i_t \times C_t'$ is expressed as the value that can be obtained C_t from the output information of the joint forgetting gate:

$$C_t = f_t * C_{t-1} + i_t * C_t'. \tag{6}$$

The calculation method is

$$\begin{aligned} O_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t * \tanh(C_C). \end{aligned} \tag{7}$$

Constructing a mathematical model for quantifying quantitative attributes, for the dimensionless value v_c' of attribute c_j, it can be expressed as

$$v_j' = \begin{cases} 1, & v_j \le v_{\min j}, \\ \dfrac{v_{\max} - v_j}{v_{\max j} - v_{\min j}}, & v_{\min j} < v_j < v_{\max j}, \\ 0, & v_j \ge v_{\max j}. \end{cases} \tag{8}$$

As a generalization of the ordinary linear model, GLM introduces a connection function in the model in order to fit some nonlinear relationships. The model can be expressed as

$$g(\xi) = g(\sigma) + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_nX_n, \tag{9}$$

where $g(\sigma)$ is the connection function, $\sigma = E(Y)$.

$$g(\xi) = a + f_1(X_1) + f_2(X_2) + \cdots + f_n(X_n). \tag{10}$$

The function of the forgetting gate is to determine the part discarded from the input information h_{t-1} and x_t and output a value between 0 and 1. The larger the value is, the more information is retained. The output of the forgetting gate is calculated as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right). \tag{11}$$

The "pumping machine" in the countryside draws on the savings of a large number of agricultural business entities such as farmers and leading enterprises but turns to urban areas to invest credit resources, resulting in a large loss of rural agricultural credit resources, thereby inhibiting the sustainable and effective development of rural agriculture. Third, government departments have strong monopolistic control of the rural financial market structure, and the policy of restricting the price of financial products has caused long-term rural credit rationing problems, which has led to a significant reduction in the credit availability of many rural households and small and medium-sized agricultural operators. The core idea of the evolution of financial market structures is as follows: first, it is manifested in a more efficient financial system, so as to build the effect of financial institution innovation and market competition; second, the financial service function of financial institutions is more stable, and the financial function changes relatively little in different periods and different political regions. Model analysis results are shown in Figure 5.

From the characteristics of the monopolistic competition type of the rural financial market structure, different types of rural financial market structure have pros and cons on agricultural growth. Existing literature research shows that from the perspective of the relationship between finance and industrial development, the two may have mutual influence or one-way influence. But in practice, the financial implementation process in many Asian countries also reflects the savings needs of poor farmers. Therefore, we need to implement savings incentives that increase the savings of the poor. The way out of poverty is to build efficient and sustainable rural financial markets, not simply to increase lending or increase savings. Agricultural credit subsidies are only available in areas where the rural financial market mechanism is seriously out of order. However, different

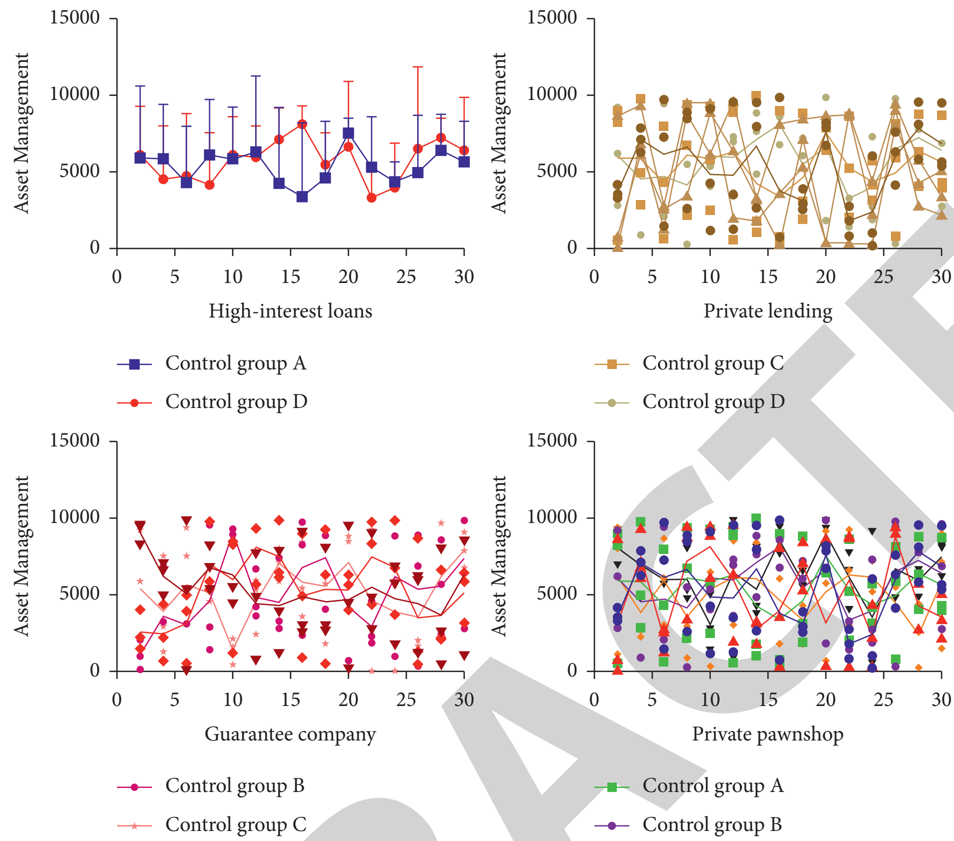


Figure 5: Liabilities of agricultural leading corporate banks in central provinces.

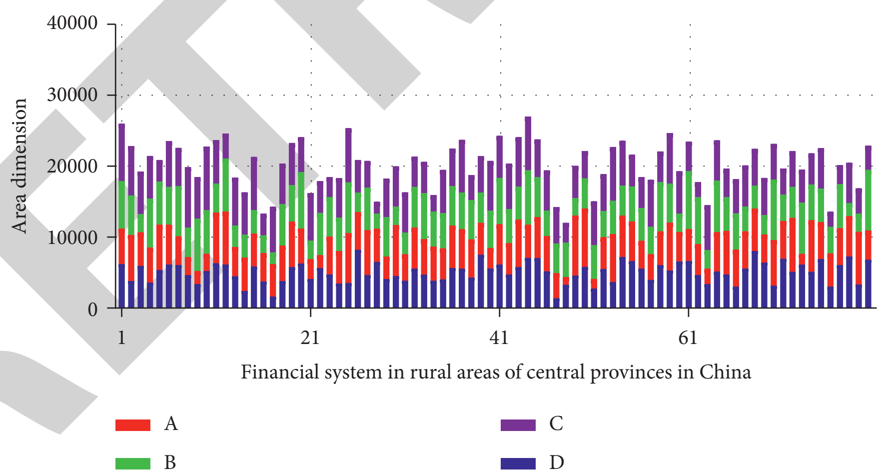


Figure 6: The influence of financial market structure on the growth of agricultural enterprise output value.

types of rural banking and other financial market structures are endogenous to the economic structure of a certain region and a certain stage of development. When investigating and analyzing the economic impact of the rural financial market structure, we must consider the rural financial market structure and the specific region in which it is located. Therefore, the matching degree of the rural financial market structure and the corresponding real economic institutions will affect the effective performance of the rural financial system's basic functions of finance. Model analysis results are shown in Figure 6.

3.2. The Influence of Financial Market Structure on Agricultural Factor Productivity. From the characteristics of the monopolistic competition type of the rural financial market structure, different types of rural financial market structure have pros and cons on agricultural growth. Existing

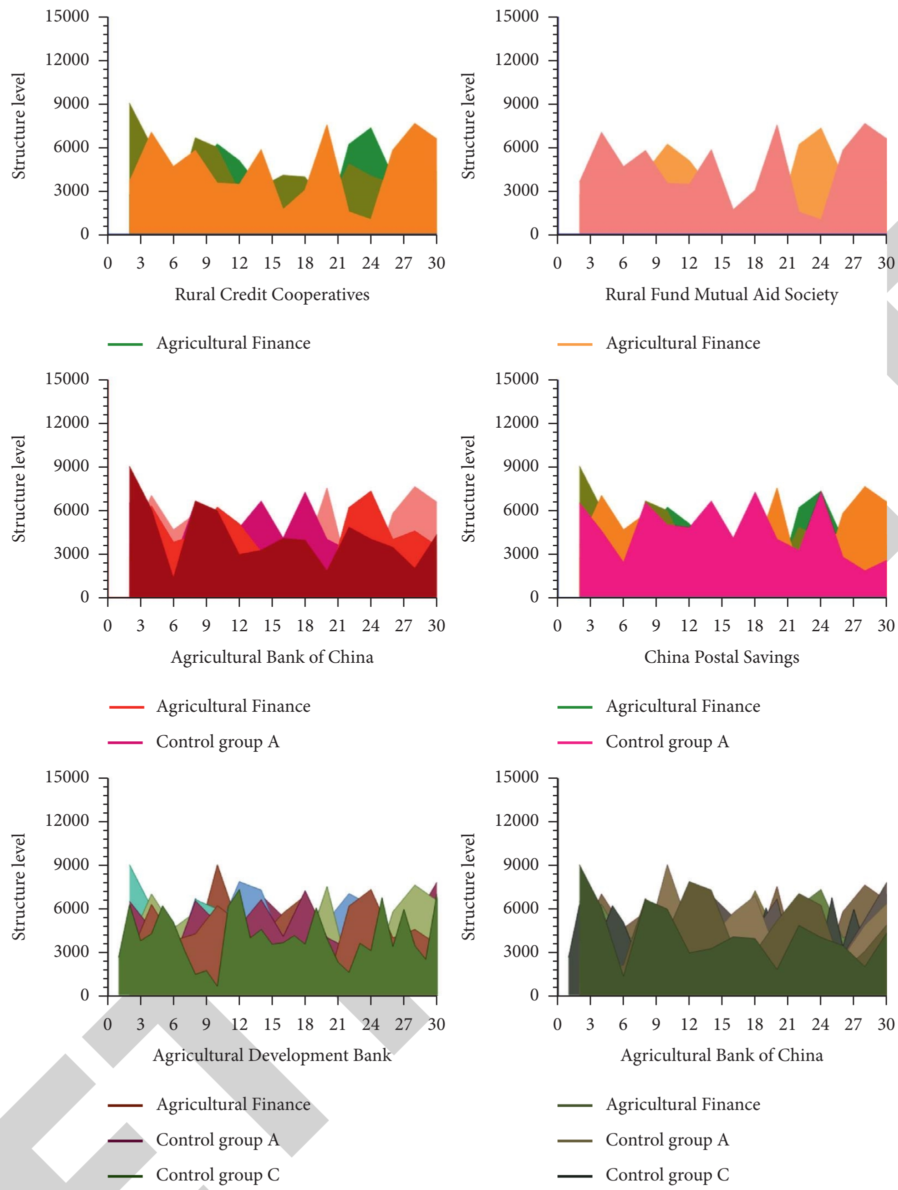


Figure 7: The influence of financial market structure on agricultural factor productivity.

literature research shows that, from the perspective of the relationship between finance and industrial development, the two may have mutual influence or one-way influence. However, different types of rural banking and other financial market structures are endogenous to the economic structure of a certain region and a certain stage of development. When investigating and analyzing the economic impact of the rural financial market structure, we must consider the rural financial market structure and the specific region in which it is located. Therefore, the matching degree of the rural financial market structure with the corresponding real economic institutions will affect the effective performance of the rural financial system's basic functions of finance. Model analysis results are shown in Figure 7.

In actual application, there is no "optimal financial structure" applicable to the economic structure of all countries (or regions). Therefore, if the specific characteristics of the real economic structure are used to judge whether the country (or region) financial structure is optimal, the evaluation result may be different. Conclusions can be drawn with the analysis of the difference in the angle of the problem. Therefore, if the specific characteristics of the real economic structure are used to judge whether the financial structure of a country (or region) is optimal. Correspondingly, the evaluation results may draw different conclusions due to different analyses of the problem angle. Although the factors affecting the financial structure of a country or region include laws, political systems, culture, production habits, etc., the characteristics of the agricultural economy's demand for rural financial services should determine a country (or region) at different stages of economic development. There are an optimal financial endowment structure and the most basic factors of its evolution. Therefore, if the factor endowment of a

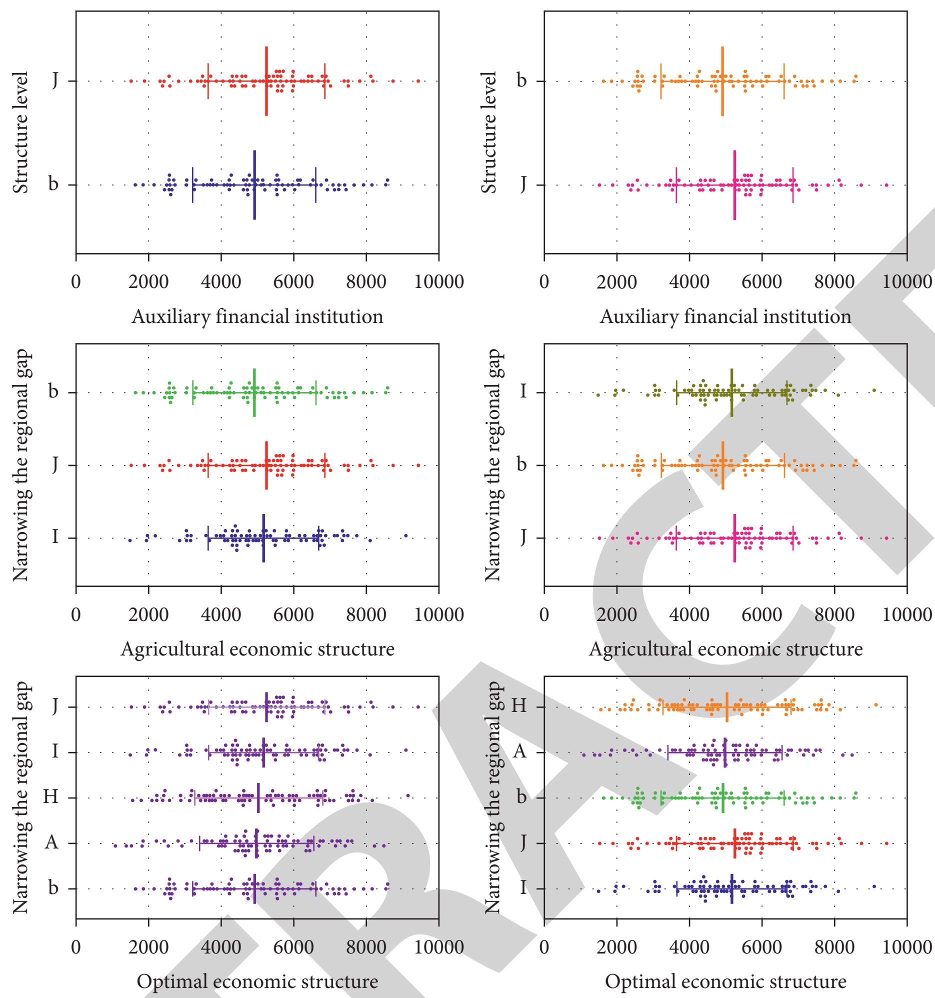


FIGURE 8: The impact of financial market structure on agricultural enterprise pollution.

country (or region) is labor-intensive, the rural financial structure should be dominated by rural financial institutions represented by regional small- and medium-sized banks that can provide services to small- and medium-sized agricultural business entities. With the improvement of the scale of business entities, technological innovation, and product quality, the proportion of large banks and securities markets that can provide financing services, diversify risks, and allocate resources for large agricultural business entities (such as leading agricultural industrialization enterprises) in the rural financial system has gradually increased. However, in practice, there is no "optimal financial structure" that applies to all countries or regions' economic structures. However, to a certain extent, the factors that affect the financial structure of a country or region include legal and political systems. At the same time, content such as culture and production habits should also be treasured. Until the advanced agricultural economies, large financial institutions and direct financing markets are in a dominant position. Model analysis results are shown in Figure 8.

As far as China is concerned, the level of development of "agriculture, rural areas, and farmers" has been continuously improved. However, the endowment of rural (or agricultural) factors determines that the main body of agricultural economic management is still mainly small- and medium-sized. Substantial changes have been made to the "small peasant" model of managing land resources with "family" as a unit. New agricultural business entities such as leading agricultural industrialization enterprises and large farmers have not replaced "small peasants" in the dominant position of the agricultural economy. However, the agricultural economy is also very important to the demand for rural financial services. The main feature of this behavior is that financial services need to adapt to the different stages of economic development of a country or region. There exist an optimal financial endowment structure and the most basic factors in its evolution. But relatively speaking, the characteristics of this factor endowment structure have regional differences in the level of evolution in different regions of China and at different stages of economic development in the same region. Therefore, in terms of agricultural growth, on the whole, the highly monopolistic rural financial market structure determined by large state-owned commercial banks or rural credit cooperatives that once dominated China's rural financial market is clearly not conducive to

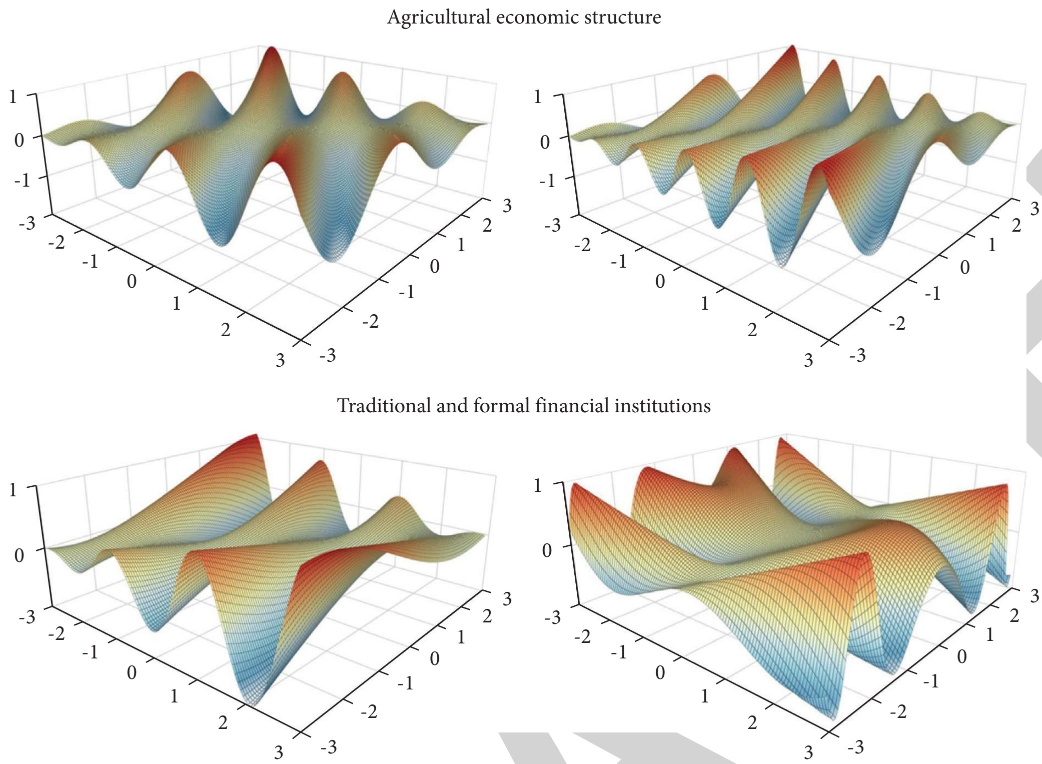


Figure 9: The measurement and factor decomposition model of agricultural total factor productivity.

agricultural growth. In terms of specific economic regions or agricultural subindustries, different regions and different subindustries are heterogeneous. If the average business scale of agricultural business entities in a region or industry is relatively large, financial institutions such as large banks should be used as rural areas. Therefore, if a region or country's production factor resources are labor-intensive, then the rural financial structure should be able to provide appropriate services for small and medium agricultural enterprises. Therefore, it is suitable for rural financial institutions represented by small- and medium-sized banks in the service area. The main body of the financial market, on the contrary, the rural financial market, should be the main body of the small- and medium-sized financial industry represented by the new rural financial institutions. At the same time, with the evolution of the factor endowments that determine the structure of the agricultural economy, the structure of the rural financial market will also be adjusted accordingly, and its impact on agricultural growth will also undergo a corresponding change. Model analysis results are shown in Figure 9.

3.3. The Impact of Financial Market Structure on Agricultural Enterprise Pollution. Agricultural total factor productivity and agricultural production input factors together constitute the source of power to drive agricultural growth in a country or region. In terms of agricultural economic growth, TFP measures the agricultural growth achieved due to agricultural technological progress, technical efficiency, scale efficiency improvement, organizational innovation, and management innovation, in addition to the input factors of agricultural production. Therefore, the total factor productivity of agriculture is as follows. Improvement plays a vital role in the process of agricultural modernization in a country or region. Agricultural development is essentially a process in which agricultural total factor productivity replaces agricultural factor inputs and its contribution continues to increase. Since the reform and opening-up, China's macro agricultural production has achieved sustained and rapid growth. Financial institutions are also very important to the development of the real economy. With the improvement of the scale of business entities, technological innovation, and product quality, it can provide financing services for large agricultural business entities (including leading agricultural industrialization enterprises). However, at present, China's workforce of the right age is showing a downward trend, the overall population growth rate is lowered, and there are a transfer of rural labor force and deployment of rural land for urbanization construction. Therefore, if China's agricultural growth still follows the extensive development model, it will be affected by existing resources. Elements and environmental constraints are as follows. The sustainable development of China's agriculture will inevitably rely on the improvement of total factor productivity (TFP) in agriculture to move toward a connotative agricultural development model. At present, the existing literature has explored the basic conditions and changing characteristics of China's agricultural total factor productivity (TFP), emphasizing the importance of TFP's impact on agriculture and the overall economic development. Model analysis results are shown in Figure 10.

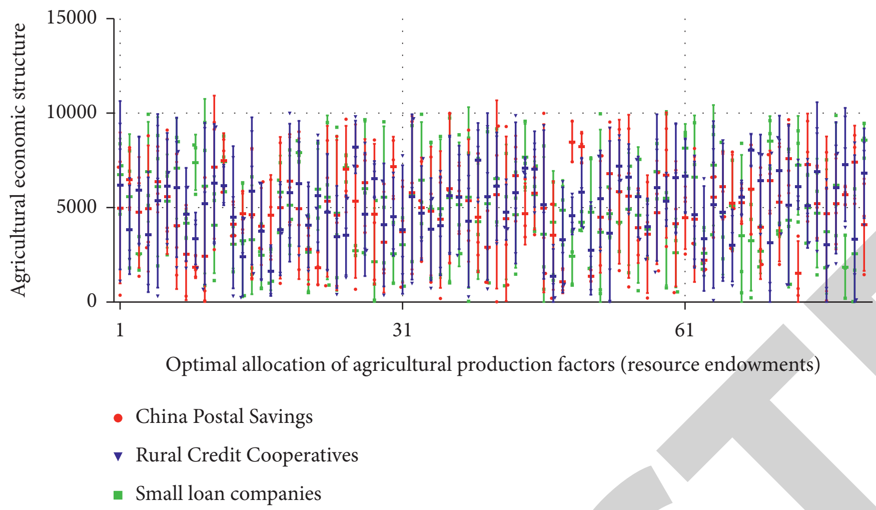


Figure 10: The regression analysis of agricultural finance to enterprise development in central China.

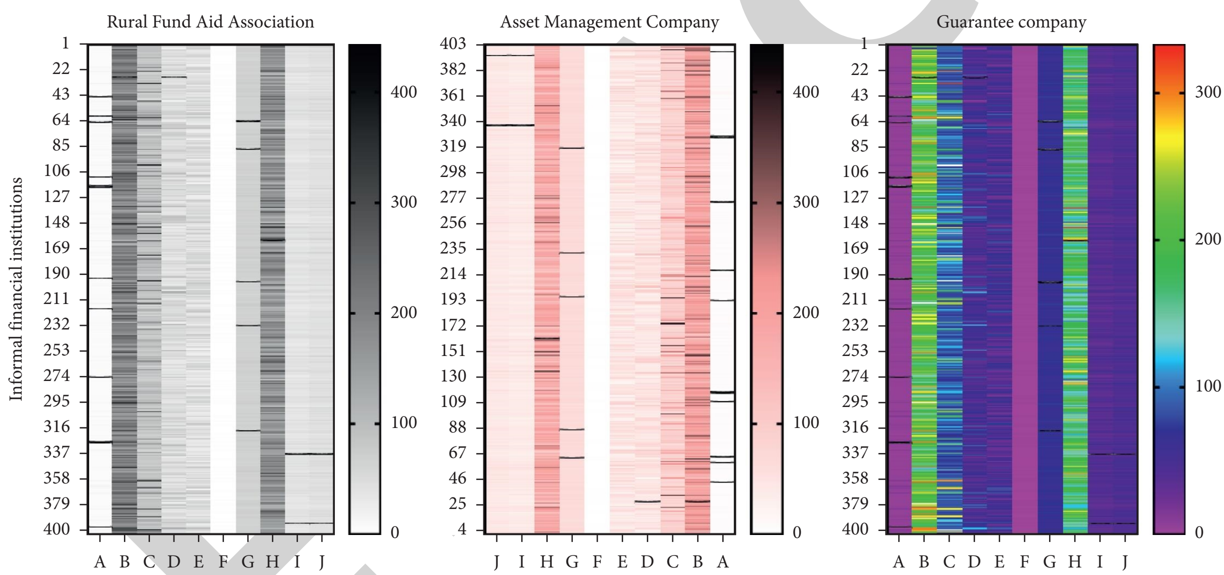


Figure 11: Financial system in rural areas of central provinces in China.

The "China Statistical Yearbook" and "China Rural Statistics Yearbook" provide more comprehensive data for measuring agricultural total factor productivity. Figure 11shows the overall time change of China's agricultural total factor productivity. It can be seen from the figure that China's agricultural total factor productivity has shown a tortuous and weak upward trend, especially in the past five years. There have been large fluctuations in the past five years. At present, China's agricultural total factor productivity lacks vitality, indicating China's agricultural development. In an important transition period, the transition from traditional agriculture to modern agriculture needs to rely on the rural financial market and other channels to provide continuous financial support for agricultural development. Model analysis results are shown in Figure 11.

Further, we analyze the decomposition of China's overall agricultural total factor productivity and its influencing factors. From the data in the table, it can be seen that the improvement of China's agricultural total factor productivity is mainly achieved through agricultural technological progress, which is basically consistent with the measurement and evaluation of most scholars. The basic growth surface of China's agricultural total factor productivity fluctuates on the "1" level, and its highest value did not exceed 1.20, indicating that China's overall agricultural total factor productivity still has a lot of room for improvement. In addition, financial institutions are also concerned with diversifying risks and allocating resources. As the proportion of large banks and securities markets in the rural financial system gradually grows, the development of China's current financial system is already similar to that of developed economies. Large financial institutions and direct financing markets dominate. From the perspective of China's agricultural total factor productivity change range, the fastest growth stage of China's agricultural total factor productivity is between 2003 and 2011. This is basically consistent with the basic growth of China's macroeconomics, indicating that the development of agriculture may also be affected by international sources. The risk assessment of order financing business is very necessary. For core companies, there are many factors that affect their risk assessment results. Generally speaking, order financing is an important way for banks and core enterprises to provide financial security for supply chain members. An important channel for enterprises to obtain financing is the bank. Corporate credit is often manifested in a variety of ways. The environmental impact of domestic agricultural development exists. Taking into account the heterogeneity of the economic development level and resource endowment conditions of different provinces, different eastern, central and western economic zones, and different agricultural and food production main functional areas, this article further analyzes the basic development characteristics of China's agricultural total factor productivity from different types of regions Describe.

4. Conclusion

The basic conclusions summarized in this article are as follows: the rural financial market structure has a systematic and profound continuous impact on China's agricultural development. The rural financial market structure that matches the agricultural industrial structure can not only provide basic financings such as continuous credit funds and risk diversification for agricultural development. Functional support, at the same time, can also provide differentiated financial services for the improvement of agricultural total factor productivity and agricultural emissions reduction and guide more construction funds to flow to the rural financial market for the optimization and upgrading of agricultural industrial structure and the optimal allocation of resources such as production factors.

Through analysis, this article draws the following basic conclusions:

(1) Since the reform and opening-up, the structure of China's rural financial market has undergone four evolutionary stages. The People's Bank of China dominates the world, the Agricultural Bank of China is in a dominant position in the rural financial market, and the rural credit cooperatives have a monopoly position, including small- and medium-sized rural financial institutions and rural banks. The structure of China's rural financial market has gone through different stages of evolution. The People's Bank of China is the top financial institution. The Agricultural Bank of China also dominates the rural financial market. Rural credit cooperatives are also in a monopoly position, including rural small- and medium-sized financial institutions and rural banks. There is a development stage of the coexistence of diversified rural financial business entities, including new rural financial organizations such as rural commercial banks, and rural financial institutions. In terms of the concentration of China's rural financial market, whether from the overall level of the country or from the provinces, cities (districts), east, middle, and west from the perspective of large economic areas or different functional areas of grain production (main grain production areas, main sales areas, production, and sales balance areas), over time, the concentration of rural financial markets is declining. New rural financial institutions such as rural commercial banks are still in the stage of coexistence and development of diversified rural financial business entities. From the perspective of the concentration of our country's rural financial market, whether it is from the overall level of the country or from the economic development of each province and city, the improvement of the financial system plays an important role. The structure of the rural financial market is still oligopolistic or highly oligopolistic. The rural financial market in economically backward areas is less competitive. From a regional perspective, there are large regional differences in the structure of China's rural financial market. There are significant regional differences no matter in the three major economic belts of the east, central, and west or in the functional areas of grain production and marketing.

(2) In terms of the overall development of China's agriculture, from the perspective of industrial growth, during the selected sample data period (2005–2016), China's agriculture has achieved worldwide agricultural development achievements. The added value and growth rate of agriculture, forestry, animal husbandry, and fishery are both a big improvement. From the perspective of the improvement of agricultural total factor productivity, China's agricultural total factor productivity has been on a slow upward process since 1997, but there have been large fluctuations between 2003 and 2011. The different functional areas of grain production (including the main grain production area, the main sales area, and the production and sales balance area) have their own main functional orientations. Over time, the concentration of rural financial markets has shown a downward trend. However, the structure of our country's rural financial market is still dominated by a single oligopoly. After that, agricultural total factor productivity even has a downward trend. There are regional differences in total factor productivity and the decomposition components of influencing factors. A phenomenon worthy of attention is that regions with a higher agricultural total factor productivity index are not regions with a higher overall economic development level. From the perspective of agricultural pollution reduction, agricultural pollutants (agricultural source total phosphorus,

total nitrogen, and ammonia nitrogen) emissions and emission intensity are still relatively high. Although the emission intensity is showing a downward trend, the issue worthy of attention is our country's major agricultural production provinces; in particular, food production areas are areas with large pollutant emissions and high emission intensity, which provide a basic reference for the implementation of differentiated agricultural pollution reduction intervention policies.

(3) Through the analysis of the mechanism of the rural financial market structure influencing China's agricultural development, it is shown that the rural financial market structure that does not match the basic economic structure (with greater market concentration) will not be conducive to the growth of China's agricultural industry. In other words, the higher concentration of rural financial markets hinders agricultural growth. The rural financial market in economically backward areas is less competitive. From the overall regional perspective, there are large regional differences in the structure of our country's rural financial market. There are significant regional differences in the three major economic belts of the east, the middle, and the west, as well as the functional areas of grain production and sales. At the same time, the higher concentration of the rural financial market is also not conducive to the improvement of agricultural total factor productivity. Agricultural pollution reduction has a certain negative impact.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Earmarked Fund for China Agriculture Industrial Research System (No. CARS-07-F-1).

References

[1] S. Banerjee and S. Venaik, "The effect of corporate political activity on MNC subsidiary legitimacy: an institutional perspective," *Management International Review*, vol. 58, no. 5, pp. 813–844, 2018.

[2] A. Edrees, H. Abdelhamed, S. W. Nho et al., "Construction and evaluation of type III secretion system mutants of the catfish pathogenEdwardsiella piscicida," *Journal of Fish Diseases*, vol. 41, no. 5, pp. 805–816, 2018.

[3] M. Guo and N. Arunkumar, "Construction of employee training program evaluation system of three exponential forecast based on sliding window," *Cluster Computing*, vol. 22, no. 3, pp. 6865–6870, 2019.

[4] K. Alexiou and J. Wiggins, "Measuring individual legitimacy perceptions: scale development and validation," *Strategic Organization*, vol. 17, no. 4, pp. 470–496, 2019.

[5] J. L. Pereira and J. O. Sá, "Process-based information systems development: taking advantage of a component-based infrastructure," *Business Systems Research Journal*, vol. 8, no. 2, pp. 71–83, 2017.

[6] S.-H. Bae and K.-J. Yoon, "Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 595–610, 2018.

[7] Z. Yang and L. S. C. Pun-Cheng, "Vehicle detection in intelligent transportation systems and its applications under varying environments: a review," *Image and Vision Computing*, vol. 69, pp. 143–154, 2018.

[8] X. Li, Y. Wang, Q.-H. Wang, Y. Liu, and X. Zhou, "Modified integral imaging reconstruction and encryption using an improved SR reconstruction algorithm," *Optics and Lasers in Engineering*, vol. 112, no. 6, pp. 162–169, 2019.

[9] S. Schnelle, J. Wang, R. Jagacinski, and H.-j. Su, "A feedforward and feedback integrated lateral and longitudinal driver model for personalized advanced driver assistance systems," *Mechatronics*, vol. 50, pp. 177–188, 2018.

[10] J. Barrena-Martinez, M. López-Fernández, and P. M. Romero-Fernández, "The link between socially responsible human resource management and intellectual capital," *Corporate Social Responsibility and Environmental Management*, vol. 26, no. 1, pp. 71–81, 2019.

[11] H. Aguinis, Y. H. Ji, and H. Joo, "Gender productivity gap among star performers in STEM and other scientific fields," *Journal of Applied Psychology*, vol. 103, no. 12, pp. 1283–1306, 2018.

[12] P. Di Girolamo, A. Behrendt, and V. Wulfmeyer, "Spaceborne profiling of atmospheric thermodynamic variables with Raman lidar: performance simulations," *Optics Express*, vol. 26, no. 7, pp. 8125–8161, 2018.

[13] F. Hegedüs, T. Bécsi, S. Aradi, and P. Gapar, "Model based trajectory planning for highly automated road vehicles," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6958–6964, 2017.

[14] A. Pipinato, P. Collin, and R. Hallmark, "Prolonging the lifetime of old steel and steel-concrete bridges: assessment procedures and retrofitting interventions," *Structural Engineering International*, vol. 29, no. 4, pp. 507–518, 2019.

[15] X. Feng, Z. Xiao, B. Zhong, J. Qiu, and Y. Dong, "Dynamic ensemble classification for credit scoring using soft probability," *Applied Soft Computing*, vol. 65, no. 4, pp. 139–151, 2018.

[16] F. Thomas and K. Christopher, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.

[17] S. Zhang, H. Wang, H. Hu, and H. Wang, "Predicting the direction of stock markets using optimized neural networks with google trends," *Neurocomputing*, vol. 285, no. 12, pp. 188–195, 2018.

[18] M. J. Mokarram, T. Niknam, J. Aghaei, M. Shafie-khah, and J. P. S. Catalao, "Hybrid optimization algorithm to solve the nonconvex multiarea economic dispatch problem," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3400–3409, 2019.

[19] Z. Khan and S. Amin, "Bottleneck model with heterogeneous information," *Transportation Research Part B: Methodological*, vol. 112, no. 1, pp. 157–190, 2018.

[20] J. M. Cairney, K. Rajan, D. Haley et al., "Mining information from atom probe data," *Ultramicroscopy*, vol. 159, no. 1, pp. 324–337, 2020.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9820128, 1 page
https://doi.org/10.1155/2023/9820128

Retraction

Retracted: Effective Bots' Detection for Online Smartphone Game Using Multilayer Perceptron Neural Networks

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Tsaur, C. H. Tseng, and C. Chen, "Effective Bots' Detection for Online Smartphone Game Using Multilayer Perceptron Neural Networks," *Security and Communication Networks*, vol. 2022, Article ID 9429475, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9429475, 10 pages
https://doi.org/10.1155/2022/9429475

Research Article

Effective Bots' Detection for Online Smartphone Game Using Multilayer Perceptron Neural Networks

Woei-Jiunn Tsaur,[1,2] **Chinyang Henry Tseng**,[2] **and Chin-Ling Chen**[3,4,5]

[1]*Computer Center, National Taipei University, New Taipei City 237303, Taiwan*
[2]*Department of Computer Science and Information Engineering, National Taipei University, New Taipei City 237303, Taiwan*
[3]*School of Information Engineering, Changchun Sci-Tech University, Changchun 130600, Jilin Province, China*
[4]*School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, Fujian, China*
[5]*Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung City 413310, Taiwan*

Correspondence should be addressed to Chin-Ling Chen; clc@cyut.edu.tw

Received 3 November 2021; Revised 30 January 2022; Accepted 3 March 2022; Published 29 March 2022

Academic Editor: Mamoun Alazab

Online smartphone game bots can cause unfair behaviors and even shorten the game's life cycle. The random forest algorithm in machine learning is a widely used solution to identify game bots through behavioral features. Although the random forest algorithm can exactly detect more definite game bot players, some players belonging to the gray area cannot be detected accurately. Therefore, this study collects players' data and extracts the features to build the multilayer perceptron, neural network model, for effectively detecting online smartphone game bots. This approach calculates each player's abnormal rate to judge game bots and is evaluated on the famous mobile online game. Based on these abnormal rates, we then use K means to cluster players and further define the gray area. In the experimental evaluation, the results demonstrate the proposed learning model has better performance, not only increasing the accuracy but also reducing the error rate as compared with the random forest model in the same players' dataset. Accordingly, the proposed learning model can detect bot players more accurately and is feasible for real online smartphone games.

1. Introduction

With the popularity of smartphones, the market of mobile online games has become more and more prosperous in recent years. The mobile online game bot is an automated program playing games autonomously instead of human users. Without any touch from a human, bots can automatically move around in a virtual world. They behave in a programmed way and repetitively do patterned actions to collect game assets. Bots can gather game wealth faster than normal players because the bot program does not get tired [1, 2]. Consequently, bots will endanger the rights of mobile online game players. For a normal player, for example, after he/she practices continuously, his/her game score should grow slowly. However, if the player suddenly increases many thousand or even ten thousand scores within a short period of time, he/she may be abnormal to use game bots. When game bots are spread, the behaviors of using bots will break the rule of the whole game and lead to the dissatisfaction of normal players. Mobile online game bots have recently become an essential security issue that must be solved to protect game assets.

Game bots' detection is one of the defensive mechanisms against game bots [3]. However, it is frequently misjudged by players as game bots because some players may spend a lot of time playing the game, which makes their score data look higher than others, or a game bot may try to imitate a normal player to avoid being detected. Many machine learning methods [1, 2, 4–8] were developed to detect game bot attacks based on behavioral features. General machine learning methods are to collect data from games' database first, and then, features will be extracted. These features are standardized to speed up the training process. After standardizing, all data are going to be separated into the training data and testing data. And the training data will be trained by some machine learning algorithms to build various models. Although these machine learning models can be used to exactly detect more definite game bot players, some players belonging to the gray area cannot be detected accurately. The existing works do not discuss the gray area of abnormal rate. These existing works use 50% to separate the normal and abnormal rates, but clusters of these two rates may overlap. So, the gray area is the cluster of detection rates between normal and abnormal rates. This work uses K means to generate the gray area cluster automatically. The rates in the gray area potentially result in most false positives and negatives. Therefore, this work will deal with these false positives and negatives in the gray area in order to increase detection accuracy.

The distribution of abnormal rates can be the clusters of high and low to predict the game bot and normal players. However, these two clusters can overlap to form a gray area such that we cannot use 50% to separate them easily. So, we need to find out those abnormal rates nearby 50%, and these rates form the gray area. Then, we can deal with these challenging rates that cause false positives and negatives. However, we require a method to find out these rates automatically instead of setting the manual border. Therefore, we adopt the K-means clustering to find out the gray area by setting $K = 3$ to cluster the rates into the normal, gray area, and abnormal clusters to generate the gray area cluster automatically. Most deep learning-based intrusion detection models do not identify and deal with the gray area directly.

In this study, we propose the multilayer perceptron (MLP) neural networks to effectively detect game bots, particularly in the gray area. As a neural network is a black-box model, it is extremely difficult for bot developers to recognize detection thresholds for neutralizing detection methods. In this study, we train the proposed deep neural network models on the famous mobile online baseball game named KANO. We find that the random forest algorithm is more frequently used in machine learning and thus choose it to compare with the proposed approach in experimental evaluation. In the experiments, we design and train four MLP models which have different optimizers, epochs, split of validation, and batch size, and we will choose the best model among them as the baseline of deep learning. We also employ the mobile online game players' dataset to train the machine learning's random forest model. According to the experimental results, the proposed best model increases the accuracy from 95.2% to 99.894% and reduces the error rate from 4.8% to 0.106%, as compared with the random forest model in the same players' dataset. Also, the proposed approach has lower FNR (False Negative Rate) and FPR (False Positive Rate) and therefore achieves better performance on detecting the gray area.

The key contributions of the study are that this paper adopts K means to generate the gray area cluster of abnormal rates automatically. These rates can result in false positives and negatives because they are nearby the 50% rate. By generating the gray area cluster, we can deal with these rates to adjust the best border to distinguish the normal and abnormal players and increase the detection accuracy. Moreover, the novelty of this work is that this work identifies the gray area by clustering the abnormal rates based on K means. The root cause of false positives and negatives comes from the gray area. Therefore, this study deals with the abnormal rate in the gray area to resolve the false positives and negatives and increase the accuracy.

The remainder of this study is organized as follows. In Section 2, the related work will be surveyed. Section 3 proposes the system model and detection approach based on the deep neural networks, including players' data feature extraction and transformation, the architecture for MLP neural networks, and gray area prediction and response. Section 4 conducts the experiments on the dataset of the mobile online game and further compares it with the random forest method in machine learning. Finally, some concluding remarks are presented in Section 5.

2. Related Works

2.1. Game Bot Attacks and Countermeasures. With the popularity of mobile online games, game bots have become a critical security issue that must be solved effectively. The behavior of using bots will break the fairness of the whole game and lead to the fact that normal players quit the game. For a normal player, his/her game score should grow slowly after he/she practices continuously. If the player suddenly increases many scores within a short period of time, he/she may be abnormal by using game bots. Based on the principle of using a game bot and the degree of endangering games, game bots are classified into three types: auxiliary, modification, and crack [1–3, 9]. The auxiliary type of game bot imitates the human's pressing button actions, automatically doing some repetitive operations. The modification type of game bot uses the third-party mobile application to inject a malicious module into the game and then modify the game's internal data. This type of game bot seriously breaks the fairness of the game. Finally, the crack type of game bot has the most influence on game vendors; it not only breaks the game's fairness but also accelerates a game to be on the decline. This type of game bot has entirely known about the game and can unscrupulously modify the game's code.

The ways of defending game bots are based on three directions [3, 8, 10]. The first solution is the client side's defense by using a mobile application, protecting the game's code from being modified to reduce the possibility of trying to crack the game. However, this method excessively drops users' game experience, as they feel disturbed during the play. Second, the information transmitted between a client and the server is encrypted to prevent game bots from eavesdropping. This solution must spend more computation time to encrypt/decrypt the transmitted information. The third method for defeating game bots is back-end protection by timely monitoring or offline data analysis to defend known or unknown game bots. However, the previously proposed models performed well in the selected game, but are hard to be utilized in other games. In addition, bot developers can neutralize detection methods if they recognize detection thresholds. Once bot developers figure out patterns of bot detection, they start to generate bots to imitate human users' behaviors [2]. For securing bot detection, it is necessary to develop an improved method, particularly in the gray area that cannot be detected accurately.

2.2. Game Bot Detection Based on Behavioral Features. Behavioral actions depict how a character performs physical activities such as moving, normal attack, or using skills [2]. Lee et al. [9] analyzed the full action sequence of players on a big data platform. Their approach was to set specific behavior sequences and apply the scoring algorithm and Naïve Bayesian algorithm to classify bots from players. In another method, Lee et al. [1] captured the similarity of character behavior to classify game bots. This method was to divide behavior sequences with a time window and embedded as a feature vector. Bots were shown they have similar behavior patterns during playtime by applying the logistic regression algorithm with self-similarity of characters. The aforementioned two methods [1, 9] demonstrated significant performance but had no sustainability. If bot developers change patterns of game bots, their detection methods are easily avoidable.

A lot of machine learning algorithms were applied to game bot defense. The most frequently used algorithms included J48, DecisionStump, HoeffdingTree, RandomForest, and REPTree [5, 11, 12]. A machine learning method is to collect data from games' databases first, and then, features will be extracted [13]. These features are standardized to speed up the training process. After standardizing, all data are going to be separated into training data and testing data. And the training data will be trained by some machine learning algorithms and built a model which is used to identify whether a player is a bot. Bernardi et al. [6] used many machine learning algorithms to detect game bots through behavioral features. They were first based on descriptive statistics to see what is the difference between a human and a bot and then utilized many machine learning algorithms to find the best algorithm that can accurately detect game bots. Kang et al. [4] detected multimodal game bots through user behavioral characteristics. They also first collected data from the game's database and then extracted users' features. They used four different machine learning algorithms to evaluate many metrics, including accuracy, precision, recall, and *F*-measure. Tao et al. [10] proposed a generalized game bot detection framework for massively multiplayer online role-playing games (MMORPGs) termed NGUARD, denoting NetEase Games' Guard. NGUARD took charge of automatically differentiating game bots from humans for MMORPGs. Xu et al. [8] employed a combination of supervised and unsupervised methods for game bot detection, where supervised methods were applied to discriminating bots and humans according to labeled data, and unsupervised methods were also applied to detecting game bots and can further help discover a new type of game bots. Park et al. [14] indicated one problem with the detection methodologies is that they do not provide rational explanations about their decisions. To resolve this problem, Park et al. [14] investigated the explainabilities of the game bot detection. They developed the XAI (Explainable Artificial Intelligence) model using a dataset from the Korean MMORPG, AION, which includes game logs of human players and game bots. Table 1 summarises the related works.

The learning models mentioned above can be used to exactly detect more definite game bot players. However, some players belonging to the gray area cannot be detected accurately. Hence, developing an improved method for effectively detecting game bots is exceedingly essential.

3. The Proposed Deep Neural Networks-Based Approach

3.1. System Model. In this study, we focus on players who use bots to modify internal data. In the game bot detection, it is hard to detect gray area players because normal or abnormal players' data look like each other in the gray area and thus reduce the accuracy of the detection system. Besides, the detection system needs more and more players' data to detect game bots; otherwise, it will only increase the false detection. This suggests the importance of players' historical data in the game detection system. Little players' data cannot increase the accuracy of the detection system.

Most back-end game bot detection approaches are to use players' data, such as players' levels or their experience values. They collect and preprocess these data from the game's database. Then, these preprocessed data will be used in machine learning or data mining to build a model or define the range. If there are new players' data that need to be detected, they can use the previously defined range to judge whether is normal or not. Different from the general approach, we use the deep learning technique [15] to build the model and utilize this model to analyze the players in the famous game named KANO which is a mobile online game of playing baseball. Figure 1 illustrates the overall architecture of the proposed game bot detection.

We devise the deep learning approach to detect game bots and gray area players based on the multiplayer perceptron (MLP) neural networks [16, 17]. MLP is a class of feedforward artificial neural networks and a universal function approximator [18]. It can be used to create a mathematical model by regression analysis. MLP has broad application situations in diverse fields such as image recognition, speech recognition, and machine translation. The proposed deep neural networks-based approach utilizes a dataset of players' historical data. The raw data are preprocessed first, including data transformation, feature extraction, and standardization, to get better data formats for model learning. Then, the preprocessed data are fed into a neural network to build a model, where we must decide how many layers and neural units the model possesses. Then, we must set each neural layer's initialization function and activation function. Finally, we set the whole model's loss function, optimizer, and metrics. After setting the model, we

Table 1: Comparison of the related works.

Author	Year	Dataset	Approach
Kang et al. [3]	2013	AION	Rule set
Chung et al. [7]	2013	Yulgang online	Multiple SVM (support vector machine)
Lee et al. [1]	2016	Lineage, AION, blade & soul	Logistic regression
Tao et al. [10]	2018	NetEase MMORPGs	ABLSTM (attention-based bidirectional long short-term memory network)
Park et al. [2]	2019	AION	LSTM (long short-term memory network)
Xu et al. [8]	2020	NetEase MMORPGs	Combination of supervised and unsupervised methods
Park et al. [14]	2021	AION	XAI (explainable artificial intelligence) model

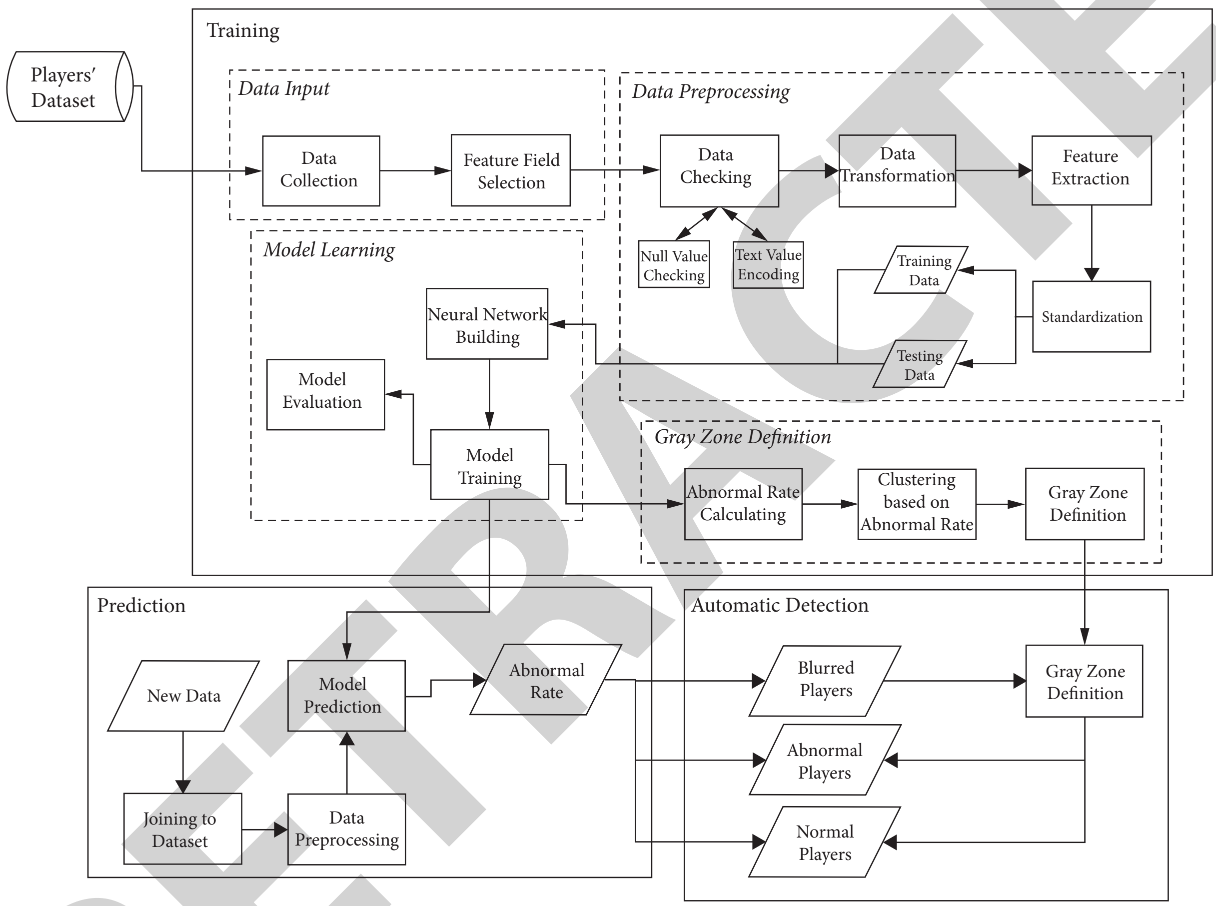


Figure 1: The system model of the proposed game bot detection.

can build the model, calculate each player's abnormal rate, cluster players based on their abnormal rate, and finally define the gray area. When there are new players whose data have been updated, the system preprocesses the raw data first and joins the players' data into the historical dataset afterward. Next, the trained model is used to predict and calculate the abnormal rate. We distinguish players into three types: normal, abnormal, and blurred player based on their abnormal rate. We present four MLP models in this study. They consist of different parameters and settings, such as kernel initialization, activation function, and epoch.

3.2. Players' Feature Extraction and Transformation. We first export players' data to a csv file named "Players.csv" from the mobile online game's database. Because we use Python to deal with the data, we choose the needed columns and transform these selected columns' data to Pandas data frame format so that we can handle the data easily. We use these fields as the training features. Table 2 shows the descriptions and names of these fields. After processing, we remove the name's column temporarily because names will not be used in the training phase and are going to be used in the prediction phase. Then, we check if there is any field being the null value and transform all data frames to arrays.

After transformation, we get a $n \times m$ matrix, which is n rows and m columns, and this is a tuple of integers indicating the size of the array in each dimension. Then, we extract the label and features from the array. Next, we are going to do standardizing because every numerical characteristic field has a different unit; for example, the unit of some players is person, and gold level's unit is level and does not have a

Table 2: Descriptions and names of player data's 9 fields.

Name	Description
Gold	The number of gold that a player holds
Gold level	The level of the number of holding gold
Player level	Player level is to depend on experience
Player experience	Player's experience
Number of players	The coach can set the number of players in the game
Player status	0: idle; 1: playing
Endurance	Endurance of a player
Fighting	Fighting of a player
Reputation	The reputation of a player

common standard. Let all values be between 0 and 1 by standardization. Standardization can raise the accuracy of the model after training. Finally, we randomly separate the processed data into training and testing data, where training and testing data are 80% and 20% of all processed data, respectively. Here, training data are used for training a model, and testing data are used for calculating the accuracy of finishing model training.

3.3. Multilayer Perceptron Neural Network Model. In this section, we design a multilayer perceptron (MLP) neural network model. MLP consists of at least one hidden layer (not including input and output layers), and it can learn not only linear function but also nonlinear function. We use backpropagation to train the model, which is a method used in an artificial neural network. Backpropagation calculates a gradient that is needed in the calculation of the weights, and it is a supervised learning method. In brief, backpropagation is just like "learning from mistakes," and it can help adjust the optimum weights.

For MLP, the multilayer network can get better convergence, but too many hidden layers just let network complexity be bigger, making more local minimum points. Cybenko [19] had proved that, in the case of a hidden layer having enough neural units, only one hidden layer can approach any continuous function. Hush and Horne [20] also presented that a neural network using two hidden layers, each layer just possessing a few neural units, can replace the neural network that uses one hidden layer and a lot of neural units. Eventually, it depends on how complex the problem is. In this study, we set the MLP as a 2-layer architecture, one input layer, two hidden layers, and one output layer [21, 22].

As for the number of neural units of the hidden layer, too few neural units may make large errors. On the contrary, although a big number of neural units can reduce deviation values, the convergence rate will slow down. Even after more than a certain number, it will increase the time of training a neural network. Taud and Mas [23] pointed out that too many neural units in the neural network generate the phenomenon of overfitting. Overfitting means overtraining; that is, the hypothesis of machine learning is too close to the training data, making the error larger when the hypothesis is going to match the testing data. Dawson and Wilby [24] presented two ways to decide the number of neural units: pruning algorithm and constructive algorithm. The pruning algorithm is a technique in machine learning. It first sets a huge number of neural units of the hidden layer to start training and then reduces the number of neural units one by one. On the contrary, a constructive algorithm sets a small number of neural units first and then adds the number of network units one by one. Because the previous processed features have 9 fields, we set the input layer as 9 neural units and concatenate the output to the hidden layer 1. The hidden layer 1 adopts a constructive algorithm and sets 20 neural units first. Finally, we find that 50 neural units possess the best effect. After we finish setting the number of neural units, we then set the initialization function. Because the essence of the training procedure of the deep learning model is to update the weight, this process requires every parameter to have its corresponding initial value, and the initialization function has a great impact on model convergence. Thus, we must set the initialization function. We set the initialization function of the hidden layer 1 as uniform distribution to initialize weight and bias:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Here, random variable x is restricted to a finite interval $[a, b]$. The activation function's main effect is to broach the nonlinearity. In the artificial neural network, if we do not use the activation function, the linear combination of the front layer input is the output of this layer (input and output still keep a linear relationship), which makes the deep artificial neural network becomes meaningless. So, we set the activation function as rectified linear units (ReLu function):

$$f(x) = \mathrm{ReLu}(x) = \max(0, x), \tag{2}$$

where x is the input to a neuron. Compared with traditional neural network activation functions, e.g., tanh and ReLu function has more advantages. ReLu function has more efficient gradient propagation, avoiding vanishing or exploding gradient problems.

Next is the hidden layer 2 which adopts the pruning algorithm. We first set 50 neural units and then find 40 neural units that have the best effect. And its initialization function and activation function are the same as hidden layer 1 [25]. In addition, because we are going to predict abnormal rates afterward, the output layer has 1 neural unit, and we set its activation function as the Sigmoid function:

$$S(t) = \frac{1}{1 + e^{-t}}, \tag{3}$$

where t is an independent variable and $S(t)$ is an exponential function. The output of the sigmoid function can range from 0 to 1.

When all functions have finished their set, we can build a sequential stack model. Figure 2 illustrates the overall MLP neural networks architecture. Before training, we have to define the training method. We first set the loss function which is used to estimate the model's discordant degree of prediction and real value. The smaller the loss function is, the greater the model's robustness is. Because the problem is binary classification, we set it as binary cross-entropy (log loss), using logistic regression, to get the following formula:

$$-\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]. \tag{4}$$

It utilizes a known sample distribution to find the parameter values that are most likely to cause this distribution (that is maximum probability). In deep learning, it is useful to use cross-entropy to obtain a better training effect. And it is common to optimize the gradient; that is, an optimizer is used to optimize every gradient descent algorithm. We set optimizer as adaptive moment estimation (Adam) function:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\sqrt{\hat{v}_t} + \varepsilon}}. \tag{5}$$

Adam optimizer is a first-order optimization algorithm that can replace the traditional random gradient descent process. It is based on training data iteration to update neural network weights. Using Adam optimizer, it can make training convergence quicker and raise the accuracy. The final parameter is the metric function. Metric is a norm of evaluating neural network performance and does not participate in the optimization process. The model is binary classification, so it usually sets the metric as an accuracy function.

Regarding the computational complexity of the proposed MLP model, we will describe it in the following. As shown in Figure 2, we define the number of neural units in the input layer, hidden layer 1, hidden layer 2, and output layer as N_i, N_1, N_2, and N_o, respectively. Because the major computational complexity in MLP involves the fully connected networks between the layers, the computational complexity is $N_i^* N_1 + N_1^* N_2 + N_2^* N_o$ [26].

3.4. Predicting Gray Area and Response. When we have finished training the model, we can further use this model to predict whether is a game bot. We write the previous data preprocessing method to a function so that when there is new or updated data needing to be detected, we can use this function to preprocess the raw data first. Then, we use the trained model to predict, and the prediction will calculate the abnormal rate to know whether a player is normal or not.

Although a player can be judged based on his/her abnormal rate, there is still some misjudgment in the gray area. In this study, the gray area means a normal player data looking abnormal or an abnormal player whose data looks normal. These two types of players make judgment go wrong easily, and therefore, we must devise a method to detect the gray area player for effectively identifying game bots.

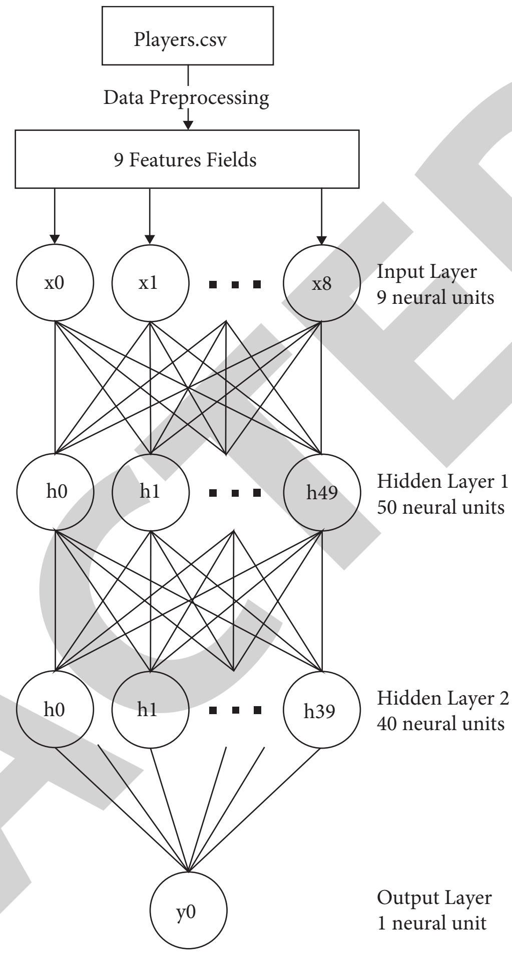


Figure 2: The proposed MLP neural networks.

We adopt the K-means clustering method [27] to cluster players based on their abnormal rates. For the abnormal rates, we want to separate them into three clusters: normal, gray area, and game bot. K means can find the best clusters with center points having minimal distances to the group members and automatically calculate the border of the cluster without manual thresholding. By K means, we can easily find out the gray area cluster between normal and game bot clusters. The rates in the gray area cluster are far away from the center points of normal and game bot clusters, so K means can help us to determine the best size of the gray area cluster and its border. The K-means clustering was presented by Hartigan and Wong [27], which is a method of vector quantization, originally from signal processing. It aims to partition n observation into K clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Through a set of observations, the general procedure is to search for a K-partition with a locally optimal within-cluster sum of squares by moving points from one cluster to another, and the following is the formula:

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x \in S_i} \left\| x - \mu_i \right\|^2. \tag{6}$$

After clustering, we obtain three clusters, including normal, gray area, and game bot clusters. To exactly define whether a player in the gray area is normal, we check each player's data to see if there is a misjudgment. Finally, we get a range to judge the player and then evaluate whether the proposed model is accurate and further compare with the random forest model in machine learning by using the following metrics [28, 29]: error rate, accuracy, precision, recall, $F1$ score, false-positive rate (FPR), and false-negative rate (FNR), where FN, FP, TN, and TP denote the number of false negatives, false positives, true negatives, and true positives, respectively:

$$\begin{aligned}
\text{Error rate} &= \frac{\text{FN} + \text{FP}}{\text{FN} + \text{FP} + \text{TN} + \text{TP}} \times 100\%,\\
\text{Accuracy} &= \frac{\text{TN} + \text{TP}}{\text{FN} + \text{FP} + \text{TN} + \text{TP}} \times 100\%,\\
\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%,\\
\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \qquad (7)\\
F1\ \text{score} &= 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%,\\
\text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{TN}} \times 100\%,\\
\text{FNR} &= \frac{\text{FN}}{\text{FN} + \text{TP}} \times 100\%.
\end{aligned}$$

4. Experimental Evaluation

4.1. Dataset Description and Preprocessing. We evaluate the proposed bot detection in the mobile online KANO game's dataset. The dataset recorded the players' data from 2019 to 2021, including normal and abnormal behaviors. We extract these three years' player data which comprise about 210,000 players (130,000 normal and 80,000 abnormal players).

The dataset also contains labels that are used to mark whether a player is normal. Labels are either 0 or 1, which means a human and a bot, respectively. Because enough data can make detection more accurate, all 210,000 data are used to train the model. The data preprocessing function is used to process the raw data, and then, the proposed MLP model will be built.

4.2. Experimental Setting and Results. In this section, we present the experimental results acquired from different MLP neural network models and a widely used random forest machine learning model with the same dataset. The training model environment is on NVIDIA GTX 1050Ti GPUs with 4 GB memory.

We set training parameters to separate the dataset into training and testing data with different ratios, where different ratios will have different results. As discussed in Section 3, Table 3 describes the specifications of models used in experiments. To simplify the expression, we name each model with a different name listed in Table 3 to explain the experimental results.

We contrast among different MLP models to choose the best model by estimating the results with the same dataset. We also contrast between the MLP model and random forest algorithm. The random forest algorithm was developed by Breiman [30], which has many merits, e.g., being able to process large input variables, and is extensively used in machine learning. Accordingly, we select the random forest method as the baseline of the general machine learning model in the experiments. We evaluate the experimental results using five metrics, including the error rate, accuracy, precision, recall, and $F1$ score. In the following, we first list the comparisons among the proposed four MLP neural network models, then conduct the comparisons between the random forest machine learning and the proposed best MLP neural network from the same KANO game players' dataset, and further make comparisons between Park et al.'s method [14] and the proposed best MLP model.

4.2.1. Comparisons among the Proposed Four MLP Models. We compare the performance of four different MLP neural network models with the same amount of data in this experiment. The differences among the four models are the optimizer, split of validation, epochs, and batch size. To strengthen the accuracy or reduce the error rate, we will increase the number of epochs and use more times of training periods to improve them. However, it still needs to consider other variables, including a split of validation or batch size; if we do not consider these variables, it may increase the training time, decrease the accuracy, and therefore raise the error rate. Table 4 demonstrates the results of the performance comparisons among these four different MLP models. We find that these four MLP neural network models reach a similar performance in the experimental results. Especially, model_3 has the best accuracy, precision, recall, $F1$ score, and the most reduced error rate. The accuracy in model_3 is extremely high, 99.894%. The $F1$ score is the combination of precision and recall. Unlike accuracy, F1 does not involve TN, true normal users, which are usually more than true abnormal users. F1 in model_3 is also very high, 96.163%. Precision shows the influences of FP against TP, and recall reflects FN. Both of them in model_3 are the best, and recall is relatively higher than precision. So, model_3 can provide extremely low FP, which is more critical than FN due to the large among of normal users, and still provide very low FN to detect abnormal users.

Moreover, Figures 3 and 4 separately illustrate accuracy and loss for training iterations of model_3 in the experiment. Chollet [31] indicated that if a proposed model has a large

Table 3: Specifications of different MLP models.

Items	Model_1	Model_2	Model_3	Model_4
Layer	2	2	2	2
Neuron	50, 40	50, 40	50, 40	50, 40
Activation function	ReLu, sigmod	ReLu, sigmod	ReLu, sigmoid	ReLu, sigmoid
Loss function	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy
Optimizer	Adam	Adadelta	Adamax	Nadam
Split of validation	0.1	0.3	0.2	0.4
Epochs	30	20	12	15
Batch_size	1000	5000	9500	7000

Table 4: Performance comparisons among different MLP models.

Metrics	Model_1	Model_2	Model_3	Model_4
Error rate	3.818%	1.382%	0.106%	1.659%
Accuracy	96.182%	98.618%	99.894%	98.341%
Precision	96.58%	97.013%	97.352%	97.116%
Recall	98.057%	97.219%	99.319%	98.127%
*F*1 score	96.123%	96.142%	96.163%	96.125%
Total training time (second)	22 s	16 s	1 s	7 s

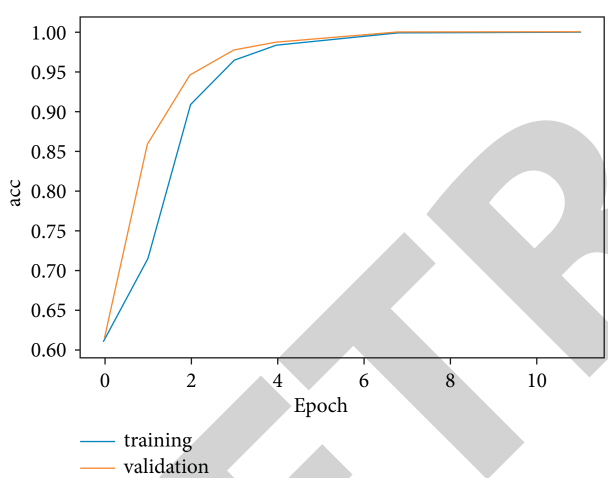


Figure 3: Accuracy for training iterations of model_3.

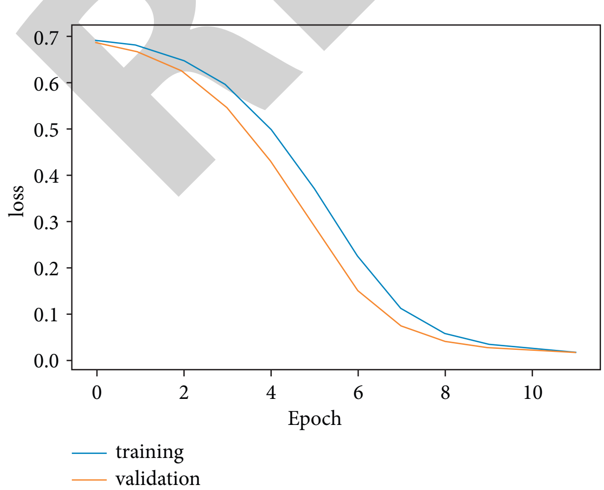


Figure 4: Loss for training iterations of model_3.

gap between its training error curve and validation error curve, it shows the proposed model may have a big problem of overfitting. In the experimental result of model_3, the gaps between the training and validation sets in accuracy and loss are both very small, which means there is not too much phenomenon of overfitting or underfitting in the training iterations.

4.2.2. Comparisons between the Random Forest Machine Learning and Proposed Best MLP Model from the Same KANO Game Players' Dataset. Here, we contrast the performance of the MLP neural network and random forest models. In the next section, we will compare the performance of detecting the gray area between the two models. Based on the above analysis, we select model_3 as the representative of various MLP neural networks and compare it with the random forest method. Table 5 shows the contrast between model_3 and the random forest model with the same dataset. The contrast result shows that the best model_3 achieves the error rate of 0.106%, accuracy of 99.894%, and *F*1 score of 99.163%, while the random forest model (with the same dataset) achieves the error rate of 4.8%, accuracy of 95.2%, and *F*1 score of 95.0%. That is, the proposed MLP neural network approach reduces the error rate from 4.8% to 0.106% and increases the accuracy from 95.2% to 99.894%.

4.2.3. Comparisons between Park et al.'s Method [14] and the Proposed Best MLP Model. In 2021, Park et al. [14] developed the random forest model and MLP model with XAI (Explainable Artificial Intelligence) modules using a dataset from the Korean MMORPG, AION, which includes game logs of human players and game bots. At first, Park et al. trained the random forest classifier and MLP classifier, respectively. The dataset contained two classes: Bot and Human, and it has been through the oversampling process due

Table 5: Performance comparisons between the MLP model_3 and random forest model.

Model name	Error rate	Accuracy	Precision	Recall	$F1$ score	Time
Model_3	0.106%	99.894%	97.352%	99.319%	96.163%	1 s
Random forest	4.8%	95.2%	95.0%	95.3%	95.0%	34 s

to the low number of data in the "Bot" class. 90% of the dataset was used to train the models, and the rest of the data were for validation. The validation accuracy of the random forest model is 88.51%, while the MLP model reached over 90.10%. Therefore, the proposed best MLP model, model_3, with an accuracy of 99.894% is superior to Park et al.'s random forest model and MLP model. Table 6 shows the accuracy comparison between Park et al.'s method and the proposed MLP model_3.

4.3. Gray Area Processing. Since model_3 has the best performance as discussed in Table 4, we use model_3 to calculate every player's abnormal rate, then cluster players based on their abnormal rate, and further define the gray area. Meanwhile, we generate data named Data_1 (20 players) and Data_2 (50 players) to detect whether the gray area is accurate or not. We also use this Data_1 and Data_2 to test the random forest method to see whether it can be accurate in detecting the gray area. In these players, each data is very close to the threshold limit value (TLV) of its field so that we can simulate players who are usually judged incorrectly in the game. We are going to use two metrics, false-positive rate (FPR) and false-negative rate (FNR), to evaluate the performance of the best MLP neural network model_3 and the random forest model. In the following, we first measure model_3 on Data_1, and the experimental result shows that FPR and FNR are both 0.0%. This means no player is misjudged, and all players are correctly detected. Then, we measure model_3 using Data_2, and the experimental result shows that FPR and FNR are 0.196% and 0.184%, respectively. On the contrary, the random forest model's FPR is 0.29% and FNR is 0.23% on Data_1 and 0.305% FPR and 0.350% FNR on Data_2. Therefore, model_3 has better detection performance, which can be more accurate to detect the gray area's players and reduce the possibilities of misjudgment.

4.4. Summary. The above experimental results show that the proposed MLP neural network model has better performance than the random forest model under the same dataset. The best MLP neural network model, model_3, increases accuracy from 95.2% to 99.894%, increases $F1$ score from 95.0% to 99.163%, and reduces error rate from 4.8% to 0.106%. In summary, our proposed model can be more accurate to detect game bots.

In terms of the gray area, the MLP neural network model can be almost fully accurate to judge the blurred players on Data_1, and it has 0.0% FPR and FNR. However, the random forest model still has a slight error; that is, it has 0.29% FPR and 0.23% FNR on Data_1. Although model_3 has a slight error on Data_2, it is a drop in the bucket as compared with the random forest model. The random forest model has a large error on judging players, and it will make players unsatisfied about it. Accordingly, model_3 is superior to the random forest model.

Table 6: Accuracy comparison between Park et al.'s method [14] and the proposed MLP model_3.

Model Name	Accuracy
Model_3	99.894%
Park et al.'s random forest model [14]	88.51%
Park et al.'s MLP model [14]	90.10%

5. Conclusions

This study proposes an effective bot detection approach for a mobile online game based on the MLP neural networks, which cannot only help enhance the performance of detecting game bots but also identify gray area players. We first collect the dataset of the mobile online game and extract its features, and then by preprocessing raw data and training the models, the proposed game bot detection approach uses four models to evaluate and further selects the best model as the baseline of the approach. The experimental results show that the proposed game bot detection model increases the accuracy from 95.2% to 99.894% and reduces the error rate from 4.8% to 0.106%, as compared with the widely used random forest algorithm in machine learning with the same players' dataset. In addition, the proposed deep neural network-based approach has better performance in detecting game bots, especially in the gray area, as compared with the widely used machine learning techniques. Furthermore, the proposed approach uses an automatic mechanism to reduce the human resources on detecting game bots and is also suitable for other games that use a back-end database to record players' data.

To extend the accomplishment of this work, we will further apply the proposed method of this work to explore the gray area in the other games. Although the other games may have different data distributions, they may have gray areas that cause false positives and negatives. This work currently utilizes the KANO game as an example to demonstrate the advantage of the proposed method. In the future, we will use the proposed method to distinguish the gray areas of the other games and check the players in the gray areas. In such a way, we will be able to show this work can also increase the game bot detection rate in the other games.

Data Availability

The data supporting this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9795431, 1 page
https://doi.org/10.1155/2022/9795431

Retraction

Retracted: Novel Multirole-Oriented Deep Learning Text Classification Model

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 22 November 2022

Security and Communication Networks has retracted the article titled "Novel Multirole-Oriented Deep Learning Text Classification Model" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] T. Luo, "Novel Multirole-Oriented Deep Learning Text Classification Model," *Security and Communication Networks*, vol. 2022, Article ID 8942841, 11 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8942841, 11 pages
https://doi.org/10.1155/2022/8942841

Research Article

Novel Multirole-Oriented Deep Learning Text Classification Model

Ting Luo

Basic Theory Department, Henan Industry and Trade Vocational College, Zhengzhou 450000, China

Correspondence should be addressed to Ting Luo; luoting@hngm.edu.cn

Received 7 January 2022; Revised 23 January 2022; Accepted 26 February 2022; Published 27 March 2022

Academic Editor: Chin-Ling Chen

In order to improve the analysis of multiple roles in novels, this article applies the deep learning text classification model to the analysis of novel roles. Moreover, in this article, the scale space formed by multiple text images of the same size is called an octave, and the text image size of adjacent groups is halved to construct a Gaussian pyramid. In text classification, this article uses the argument and amplitude information to form a direction gradient histogram and takes the argument corresponding to the largest peak as a main direction of the key point. Finally, this article constructs an intelligent analysis model. The research results show that the deep learning text classification model for multiple roles in novels proposed in this article has good effects on role analysis and text classification. For film and television scripts, the classification analysis of analysis texts and the contrast creation have very good application help.

1. Introduction

The analysis of the character's personality in the novel helps us to understand the psychology of the character in the novel and the real life it reflects more accurately and deeply. At present, there are two main ideas for the character analysis of novel characters in academic circles. One is the qualitative study of literature and art. According to the researcher's humanistic quality and the reading experience of the novel text, the personality of one or more characters in the novel is summarized into the main aspects from the micro-perspective. Then, the character characteristics of these aspects are corroborated by the relevant descriptions in the text [1], and the characteristics of the characters in the novels written by the same author are further summarized at the macro level [2]. The other is from the perspective of linguistics, by selecting several dialogues in the novel for pragmatic principles, turn-taking or other conversational analysis, or using a corpus to count the words used in the characters' language, and get the words with higher frequency as keywords, and analyze the personality of the characters through the keywords [3]. At present, there are few studies that go deep into the initial character personality or personality traits (the first level) to discuss the role.

Personality is a person's stable attitude towards reality and habitual behavior, and it is a significant tendency of an individual's behavior. From the perspective of psychometrics, compared with other mental states that are relatively easy to change, personality is a relatively stable personality trait, which has the advantages of being comparable, descriptive, and cross-contextual. Therefore, personality or personality traits can be regarded as the behavioral expression or mode of individual psychological activities. Psychological analysis of the characters in the novel from the perspective of personality can help to have a more accurate and profound understanding of the characters in the novel. Describing literary characters through their personalities can intuitively see their most prominent psychological characteristics and compare the psychological differences of different characters. The research field of personality traits includes a variety of theoretical models, among which the "big-fivepersonalitytrait model" (hereinafter referred to as "big-fivepersonality") is the most valued by today's academic circles.

This paper applies deep learning text classification technology to the analysis of multiple roles in novels and builds an intelligent analysis model, which provides a theoretical reference for intelligent literature.

The current research on artificial intelligence literary creation has the shortcomings of a single research angle and not deep theoretical foundation, so the research motivation of this paper is to improve the effect of artificial intelligence literature analysis.

In order to improve the effect of multirole analysis in novels, this article applies the deep learning text classification model to the analysis of novel characters so as to realize the analysis effect of intelligent novel characters.

The main research structure of this article is as follows: the introduction part analyzes the research status of multiple roles in novels and draws out the research content of this paper. The second part is the research algorithm of this article. The third part mainly studies the text classification model and proposes an improved algorithm as the basis for the model construction in this article. The fourth part combines the deep learning text classification method to construct a novel multirole analysis model. The reliability of the method in this article is verified by analysis. The conclusion part summarizes the research content of this article.

2. Related Work

Character modeling is a representation of characters that are automatically generated from the text of literary works. Dharmawan extracted the salient characteristics of characters from the text of the story, formed a description of the characters, and then generated a summary of the story [4]. Finck used topic models to learn character types from movie plot summaries, and each character type is represented from three aspects: agent, patient, and attribute according to the dependency relationship, and each aspect is represented. It is the distribution of a series of hidden topics, and each hidden topic is expressed as the distribution of words [5]. He used topic models to learn character types from novel texts. Each character type is based on four aspects: agent, patient, possessive, and predictive. Representation, each aspect is directly expressed as the distribution of words, and the factor of the novel author is added when generating words. Character profiling is an explicit way of modeling characters. It is an automatic classification problem to predict the attributes of characters, such as gender, age, and personality, from the text of literary works [6]. Hudson predicted whether the character of the novel is introverted or extroverted, using the SVM classifier, using the three characteristics of the character's words, the character's actions, the character's adjective or adverbial description, and adding a variety of vocabulary resources, such as WordNet, VerbNet, LIWC, and word vector. The results show that using characters' words as features are not as effective as the latter two. In terms of automatic classification of character relationships [7], Kamble divided the relationship between the characters in the novel into three categories: social, professional, and familial, and each relationship is further subdivided; according to the closeness of the relationship, it is divided into positive), neutral, and negative, and 109 English novels are marked with character relationships [8]. Drawing lessons from the idea of "the enemy of the enemy is a friend," Lunyachek introduced structural features other than textual features when predicting the friendly or opposite relationship between characters based on the movie plot summary, that is, the relationship between the target character and other characters. This is useful for judging the relationship between target characters and modeling changes in character relationships. The length of the novel is longer. With the development of the plot, the relationship between the characters will often change. At this time, it is not necessary to explicitly define the relationship type, and the distribution of words can be used to express the relationship. The change of relationship is a sequence problem [9]. Maruthu proposed a character relationship model (relationship modeling network, RMN) based on recurrent neural network (RNN) and dictionary learning. The relationship type is represented by a vector, and the set of words most similar to the relationship type vector is used. To describe the relationship type [10]. Matulionyte used Hidden Markov Model with Gaussian emission to model changes in character relationships, taking into account various features such as dependent verbs, bag-of-words, and semantic frames [11]. McSherry extended RMN and, at the same time, modeled character changes and character relationship changes. The two changes were combined to express novels, calculate the similarity between novels, and then realize novel recommendations. For character network extraction and analysis [12], extract the characters and relationships in the novel, construct a network of characters, and discover or verify the characteristics of the characters in the novel and their social environment by studying the nature of the network. The nodes of the network are the characters, and the edges are the relationships between the characters. The definition of relationship can be a cooccurrence relationship between characters [13]; that is, the characters appear together in a certain scale of context, such as a sentence, a paragraph, or a chapter; it can also be an event relationship; that is, the characters participate in the same event. [14]. It can also be a dialogue relationship; that is, there is a dialogue interaction between two people [15]. Moreover, it can be some explicitly defined relationship, such as family relationship, and work relationship. The above research work is based on basic tasks such as word segmentation, named entity recognition, and interlocutor recognition in literary texts. At present, these tasks perform well in news texts, while in literary texts, they are faced with domain migration. The huge challenge of this has gradually attracted the attention of more researchers [16].

To sum up, the current research on artificial intelligence literary creation has a single research angle and weak theoretical foundation. Therefore, based on the existing research results of the academic circle, I will clarify the research object from the source-artificial intelligence on the one hand. Connotation and development context, based on this basis, make a basis and targeted evaluation of the current situation of artificial intelligence literary creation and, on the other hand, strengthen theoretical interpretation, give play to the guiding value of professional theory, and analyze specific cases of machine creation. They rise to the height of theory, not only based on the present and linking with the reality, but also keep an eye on the future and keep up with the trend of the times.

3. Intelligent Text Classification Model

The feature extraction and description of text images is the most important part of the passive forensics algorithm of digital text image copy-paste tampering.

Firstly, the text feature extraction algorithm is analyzed, and the pair analysis is carried out.

SIFT can extract a feature with rotation and scale invariance, which is called SIFT feature, which has strong robustness to noise and brightness transformation. SIFT includes two steps: key point detection and key point feature description. Among them, key point detection includes scale space extreme value detection and key point positioning, and key point feature description includes direction assignment and key point description.

First, the text image and the Gaussian kernel function are convolved to construct a scale space, which is represented by $L(x, y, \sigma)$, as shown in the following formula [17]:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y). \tag{1}$$

In the formula, $*$ means convolution in x- and y-directions and $I(x, y)$ means text image. $G(x, y, \sigma) = 1/2\pi\sigma^2 e^{-(x^2+y^2)/2\sigma^2}$, which is a Gaussian kernel function with variable scale.

The scale space formed by multiple text images of the same size is called an octave, and the text image size of adjacent groups is halved. As a result, a Gaussian pyramid is constructed. In each group of text images, two adjacent text images are subtracted, as shown in formula (2). In this way, the text image $D(x, y, \sigma)$ can be obtained to form a DoG (Difference of Gaussian), which is often referred to as a DoG pyramid. Then, the local extreme points are searched in the difference pyramid to determine candidate key points in the scale space and location space:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned} \tag{2}$$

In the formula, k represents a constant.

Since the candidate key points are selected from text images of different sizes, the coordinates need to be accurately fitted. If it is assumed that the candidate extreme point $\mathbf{x}_0 = (x_0, y_0, \sigma_0)^T$, this article uses this point as the origin to perform the second-order Taylor expansion of $D(x, y, \sigma)$:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}. \tag{3}$$

In the formula, $\mathbf{x}$ represents the offset relative to $\mathbf{x}_0$. In this article, formula (3) is used to obtain the derivative of the independent variable $\mathbf{x}$ and $\widehat{\mathbf{x}}$ is the location of the extreme point. If the value of $\widehat{\mathbf{x}}$ in any dimension exceeds 0.5, which means that the extreme point is closer to the neighboring point of point $\mathbf{x}_0$, so the neighboring point is used as the origin to revalue [18]. The algorithm repeats this process until the value of $\widehat{\mathbf{x}}$ in all dimensions does not exceed 0.5. Substituting the offset $\widehat{\mathbf{x}}$ into formula (3), the extreme value $D(\widehat{\mathbf{x}})$ of the extreme point in the difference pyramid is obtained [19]:

$$D(\widehat{\mathbf{x}}) = D + \frac{1}{2}\frac{\partial D^T}{\partial \mathbf{x}}\widehat{\mathbf{x}}. \tag{4}$$

$|D(\widehat{\mathbf{x}})|$ is used as an index to measure the contrast of extreme points, and it is considered that the extreme points of $|D(\widehat{\mathbf{x}})| < T_c$ are susceptible to noise and are directly discarded. Lowe recommends setting $T_c = 0.03$ (assuming that the pixel value of the text image is between 0 and 1).

Since $D(x, y, \sigma)$ responds very strongly to the edge, the extreme points on the edge are very susceptible to noise. The Gaussian difference function will produce a larger principal curvature at the edge, and the principal curvature can be calculated by the Hessian matrix shown in the following:

$$\mathbf{H} = \begin{bmatrix} D_{\mathrm{xx}} & D_{\mathrm{xy}} \\ D_{\mathrm{xy}} & D_{\mathrm{yy}} \end{bmatrix}. \tag{5}$$

The eigenvalues of the Hessian matrix are proportional to the principal curvature. In order to avoid calculating eigenvalues directly, formula (7) is used to measure the eigenvalues [20].

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r+1)^2}{r}. \tag{6}$$

In the formula, $\mathrm{Tr}(\mathbf{H})$ is the trace of matrix $\mathbf{H}$, $\mathrm{Det}(\mathbf{H})$ is the determinant of matrix $\mathrm{Det}(\mathbf{H})$, α is the larger eigenvalue of $\mathbf{H}$, and β is the smaller eigenvalue of matrix $\mathbf{H}$, $r = \alpha/\beta$. Finally, for each candidate key point, this article judges whether its matrix $\mathbf{H}$ satisfies formula (7):

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(T_r + 1)^2}{T_r}. \tag{7}$$

In the formula, T_r is the edge threshold, and the recommended value is 10. Eliminate the extreme points that do not satisfy formula (7), and the remaining points are the key points extracted according to SIFT.

According to the local nature of the text image, the direction of each key point is assigned so that the feature descriptor has rotation invariance. For each text image $L(x, y)$, this article uses the key point as the center and the radius $3 \times 1.5\sigma$ of the circular area of the text image gradient to calculate the amplitude $\theta(x, y)$ and amplitude $m(x, y)$ [21]:

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \\ \theta(x, y) &= \tan^{-1}\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}. \end{aligned} \tag{8}$$

In this article, the directional gradient histogram is composed of the angle and magnitude information. The argument corresponding to the maximum peak value is taken as one main direction of the key point, and the argument corresponding to the remaining peak value 80% higher than the maximum peak value is taken as the remaining main direction of the key point. Therefore, the coordinate positions and scales of the key points generated by this area are exactly the same, but the main directions are different. At this point, the coordinates, scale, and direction of the text image can be assigned to each key point.

Hu moment is a commonly used feature in the CMFD algorithm based on text image segmentation. This feature is called Hu moment, which is also often called invariant moment. Below, ordinary moments are introduced first.

The $p+q$-order moment of a density distribution function $\rho(x, y)$ is defined as shown in the following formula:

$$m_{\mathrm{pq}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q \rho(x, y)\mathrm{d}x\mathrm{d}y,\quad p, q = 0, 1, 2, \ldots \tag{9}$$

For the discrete text image $f(x, y)$, the definition of the $p+q$-th order moment is rewritten as the following formula:

$$m_{\mathrm{pq}} = \sum_{x=1}^{M}\sum_{y=1}^{N} x^p y^q f(x, y). \tag{10}$$

In the formula, M represents the number of pixels in the x-direction of the text image, and N represents the number of pixels in the y-direction of the text image. From formula (11), we can construct 0-order m_{00}, 1st-order moments m_{01}, m_{10}, etc. The physical meaning of the 0th moment is the mass of the target area, and the 1st moment represents the center of mass of the target area. Constructed invariant moments are based on these ordinary moments. In order to obtain translation invariance, the central moment shown in formula (11) is constructed:

$$\mu_{\mathrm{pq}} = \sum_{x=1}^{M}\sum_{y=1}^{N} (x - x_0)^p (y - y_0)^q f(x, y). \tag{11}$$

In the formula, $x_0 = m_{10}/m_{00}, y_0 = m_{01}/m_{00}$. This represents the centroid position of the text image. In order to obtain scale invariance, the central moment is normalized according to the following formula:

$$\eta_{\mathrm{pq}} = \frac{\mu_{\mathrm{pq}}}{\mu_{00}^{r}}, p + q = 2, 3, \ldots, \tag{12}$$

where η_{pq} is the normalized central moment of order v, $r = p + q/2 + 1$. Using the second- and third-order normalized central moments, seven kinds of Hu moments with translation, rotation, and scale invariance can be constructed, of which four low-order Hu moments are shown in the following:

$$\begin{aligned}
\Phi_1 &= \eta_{20} + \eta_{02},\\
\Phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2,\\
\Phi_3 &= (\eta_{20} - 3\eta_{12})^2 + 3(\eta_{21} - \eta_{03})^2,\\
\Phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2.
\end{aligned} \tag{13}$$

In the formula, η_{20}, η_{02}, and η_{11} are the second-order normalized central moments, and $\eta_{30}, \eta_{03}, \eta_{12}$ are the third-order normalized central moments.

Next, the intelligent text feature analysis algorithm is synthesized, and the process of this algorithm is proposed.

Through analysis, different algorithms are selected in each link. This article combines them according to a certain process and can realize the improved algorithm of passive forensics based on SIFT copy-paste tampering of text and images. The concrete realization procedure is shown in Figure 1.

It can be seen from Figure 1 that the implementation process of the improved algorithm based on SIFT is as follows. First, the algorithm uses an improved feature extraction method to obtain a large number of SIFT features and improves the heavy-tailed distribution by squaring the features. Secondly, in order to improve the speed of nearest neighbor search, the algorithm uses a brute force search algorithm based on matrix operations to find the nearest neighbors of feature points. Third, in order to match as many similar key points as possible, the algorithm uses a three-step feature matching method to match similar features. Finally, in order to effectively filter out mismatches, the algorithm first uses cross-checking to initially filter and remove duplicate point pairs, then uses hierarchical clustering to filter out isolated points, and then uses the RANSAC-based affine transformation matrix estimation method to filter out the excluded points.

The steps of the brute force search algorithm based on matrix operations are as follows:

(1) This algorithm constructs matrices **F** and **C**. **F** is an $n \times 128$-dimensional matrix. Each row represents the feature of a key point. There are n key points in total, and $\mathbf{C} = \mathbf{F}$. Therefore, calculating the distance between the features is transformed into calculating the Euclidean distance between each row of matrix **F** and each row of matrix **C**. The symbol dist represents the Euclidean distance between the rows of matrix **F** and matrix **C**, where the element $\mathrm{dist}(i, j)$ in the matrix dist represents the Euclidean distance between the i row $\mathbf{F}_i$ of the matrix **F** and the j row $\mathbf{C}_j$ of the matrix **C**. The calculation formula of dist is shown in the following formula:

$$\mathbf{dist} = \sqrt{\mathbf{A} + \mathbf{B} - 2\mathbf{D}}. \tag{14}$$

In the formula,

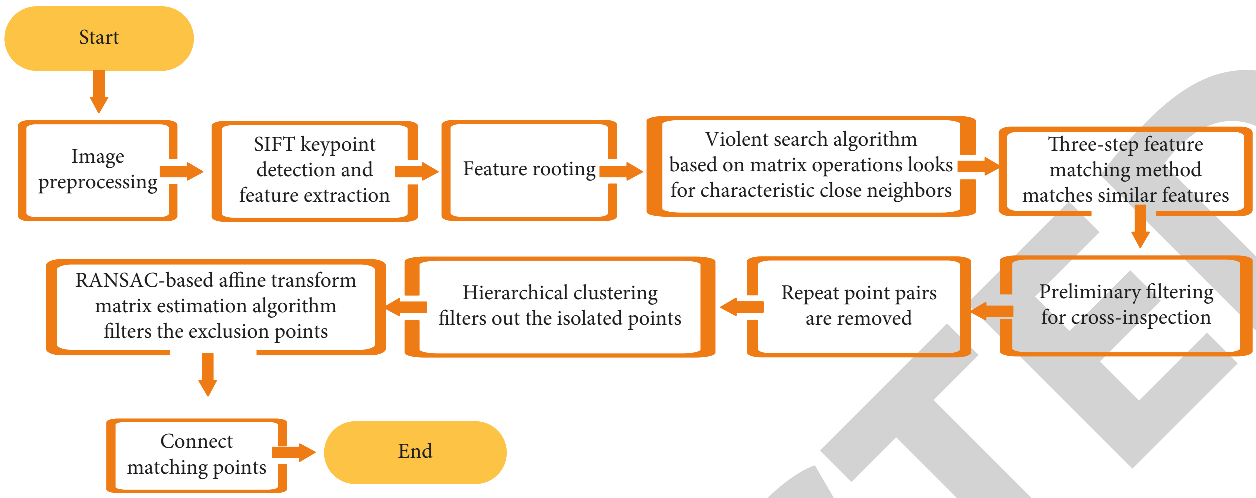


Figure 1: SIFT-based text image copy-paste tampering passive forensics improved algorithm flowchart.

$$\mathbf{A} = \begin{bmatrix} \|\mathbf{F}_1\|^2 & \|\mathbf{F}_1\|^2 & \cdots & \|\mathbf{F}_1\|^2 \\ \|\mathbf{F}_2\|^2 & \|\mathbf{F}_2\|^2 & \cdots & \|\mathbf{F}_2\|^2 \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{F}_n\|^2 & \|\mathbf{F}_n\|^2 & \cdots & \|\mathbf{F}_n\|^2 \end{bmatrix},$$
$$\mathbf{B} = \begin{bmatrix} \|\mathbf{C}_1\|^2 & \|\mathbf{C}_2\|^2 & \cdots & \|\mathbf{C}_n\|^2 \\ \|\mathbf{C}_1\|^2 & \|\mathbf{C}_2\|^2 & \cdots & \|\mathbf{C}_n\|^2 \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{C}_1\|^2 & \|\mathbf{C}_2\|^2 & \cdots & \|\mathbf{C}_n\|^2 \end{bmatrix}, \tag{15}$$
$$\mathbf{D} = \mathbf{F}\mathbf{C}^T;.$$

(2) After the algorithm obtains the distance matrix dist, it is noticed that the diagonal elements are all 0, which indicates the distance between a feature and itself. In order to facilitate subsequent calculations, the diagonal element value is set to −1.

(3) The algorithm uses the quick sort method to sort the distances in ascending order and records the index corresponding to each distance. In the sorting result, the point with the smallest distance from x_i is x_i itself, which is removed.

Through the above steps, the nearest neighbors of each key point in the feature space can be obtained.

Taking the key point x_i as an example, the steps of the three-step feature matching method are explained as follows:

(1) The algorithm improves the g2NN criterion. Through the brute force search algorithm based on matrix operation, the matching degree vector $\mathbf{D} = \{d_1, d_2, \ldots, d_{n-1}\}$ of $\mathbf{x}_i$ is obtained, and the ratio vector $\mathbf{r} = \{d_1/d_2, d_2/d_3, \ldots, d_{n-1}/d_n\}$ is calculated. Then, the algorithm finds the last element d_k/d_{k+1} in the vector $\mathbf{r}$ that is less than the threshold T and thinks that the point $\mathbf{P}_k$ and all the key points $\{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_{k-1}\}$ before it may be the key points that match the $\mathbf{x}_i$ point.

(2) Considering that in the text image, the key point features of adjacent positions may be very similar. Therefore, the distance between point $\mathbf{x}_i$ and point $\{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_k\}$ in the text image is required to be greater than $|V_s|$, and $|V_s|$ is the movement vector threshold.

(3) In order to prevent the occurrence of a large distance between features and a sudden change in the ratio, the absolute threshold T_{abs} is used to filter the matching points. In order to automatically select the appropriate T_{abs}, referring to the two-stage feature matching method proposed by Jin Guonian, the average distance between the matching features that meet the 2 NN detection is used as the threshold T_{abs}. In particular, when choosing the matching features detected by Laughter 2 NN, a simple mismatch filtering was performed using cross-checking.

Through these three steps, the matching points of point $\mathbf{x}_i$ can be screened out. For the remaining key points, the same steps can be used to find the matching point of each key point.

The basic principle of the cross-checking algorithm is as follows: if point x_i exists in the matching point of point x_j, then point x_j should also exist in the matching point of point $\mathbf{x}_i$. For each key point, a check is carried out, and preliminary filtering can be carried out quickly.

Condensed hierarchical clustering first divides each point into a cluster individually, then continuously merges the closest clusters, and finally merges into a cluster. The hierarchical structure of the entire cluster is similar to an inverted tree.

There are many ways to calculate the distance between clusters. Experiments have found that in CMFD, using the Ward method to calculate the distance is the best. At this time, the clustering algorithm is also called the Ward-linkage algorithm. The distance calculation formula of the Ward method is shown in the following formula:

$$d(u,v)=\sqrt{\frac{|v|+|s|}{T}d(v,s)^2+\frac{|v|+|t|}{T}d(v,t)^2-\frac{v}{T}d(s,t)^2}. \quad (16)$$

In the formula: u, v, s, t are all clusters, and u is a new cluster formed by the fusion of s and t, $T=|v|+|s|+|t|$, and the symbol $|\bullet|$ represents the number of elements in the set "$\bullet$."

When the key points are clustered according to the spatial distance, the "pruning" operation at the height can form m clusters, which are represented by the set $\{c_1,c_2,\ldots,c_m\}$. c_i is selected from the set in the order of $i=1,2,\ldots,m-1$, and for each c_i, a pairing attempt is made with c_j, where $j=i+1,i+2,\ldots,m$.

The pairing principle is as follows: if there is a matching point of the point in c_j in c_i, then c_i and c_j are paired successfully, and the matching point pair in c_i and c_j can form a cluster pair C_k. Among them, $C_k=\{\{(x_{k_1},y_{k_1}),(x'_{k_1},y'_{k_1})\},\{(x_{k_2},y_{k_2}),(x'_{k_2},y'_{k_2})\},\ldots\{(x_{k_n},y_{k_n}),(x'_{k_n},y'_{k_n})\}\}$, $\{(x_{k_i},y_{k_i}),(x'_{k_i},y'_{k_i})\}$, which represents a pair of matching points, and $i=1,2,\ldots n$. This algorithm counts the cluster pairs generated according to the set $\{c_1,c_2,\ldots,c_m\}$ to form the cluster pair set $\{C_1,C_2,\ldots,C_s\}$. For any cluster pair C_k in the set, $\mathrm{card}(C_k)\geq 3$ is required, where the symbol $\mathrm{card}(\bullet)$ refers to the number of elements in the set $\bullet$, and the number of elements greater than 3 takes into account the requirements of affine homography matrix estimation. The algorithm renumbers the cluster pairs that meet the requirements of the number of elements to form a set $\{C_1,C_2,\ldots,C_l\}$. The set can be divided into $2l$ clusters, which can be regarded as $2l$ different regions.

According to the generalized text image copy-paste tampering model, the text image copy-paste transformation method can finally be summarized as an affine transformation. The affine transformation form is given in the formula, which is rewritten as a matrix form, as shown in the following formula:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix}=\begin{bmatrix} m_{11} & m_{12} & \Delta x \\ m_{21} & m_{22} & \Delta y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}\longrightarrow \mathbf{X}'=\mathbf{HX}. \quad (17)$$

In the formula, $m_{11},m_{12},m_{21},m_{22}$ represent linear transformation parameters, $\Delta x,\Delta y$ represent translation parameters, x and y represent the key point coordinates of the copy, and x' and y' represent the coordinates of the key point of the paste. $\mathbf{H}$ represents the affine transformation matrix, which is also called affine homography.

Generally, to solve $\mathbf{H}$, at least three sets of matching point pairs are required, and three sets of matching point pairs are taken out of cluster pairs $C_k, k=1,2L,l$. Assuming that under ideal conditions, these point pairs use the same transformation matrix, and the three points in a cluster are not collinear, the affine homography matrix can be calculated by $\mathbf{H}=\mathbf{X}'\mathbf{X}^T(\mathbf{XX}^T)^{-1}$. In the text image, the coordinates are discrete, and there will be problems such as mismatched point pairs and point collinearity. Therefore, it is necessary to use the least-squares method to estimate matrix $\mathbf{H}$ and combine the RANSAC algorithm to divide the matching point pairs into inliers and the mismatched point pairs into outliers, and through repeated iterations, the real $\mathbf{H}$ is constantly approached.

$$L(\mathbf{H})=\sum_{i=1}^{kn}\left\|\mathbf{X}'_i-\mathbf{HX}_i\right\|^2. \quad (18)$$

For cluster pair $C_k, k=1$, we set the current optimal affine homography matrix as $\mathbf{H}_b=$ None, and the number of interior points of C_k under matrix $\mathbf{H}_b$ is $M_b=0$. The steps of the RANSAC-based affine transformation matrix estimation algorithm are as follows:

(1) The algorithm randomly selects three samples from $C_k=\{\{(x_{k_1},y_{k_1}),(x'_{k_1},y'_{k_1})\},\{(x_{k_2},y_{k_2}),(x'_{k_2},y'_{k_2})\},\ldots\{(x_{k_n},y_{k_n}),(x'_{k_n},y'_{k_n})\}\}$ to form $\{\{(x_1,y_1),(x'_1,y'_1)\},\{(x_2,y_2),(x'_2,y'_2)\},\ldots\{(x_m,y_m),(x'_m,y'_m)\}\}$ and checks whether the data are valid. If the data are invalid (less than three linearly independent points), the algorithm jumps out of this iteration. Otherwise, the algorithm uses the least-squares method to calculate the affine homography matrix $\mathbf{H}$.

(2) This algorithm constructs $X_i=(x_i,x_j),\quad j=1,2,L$. k_n is substituted $\widetilde{\mathbf{X}}'_i$ into formula (18) to obtain $\widetilde{\mathbf{X}}'_i$. According to formula (19), the inner and outer points are divided. The point pairs satisfying the formula (19) are divided into inner point pairs $X_i=(x_i,x_j),\quad j=1,2,L$, and the remaining points of M are divided into outer point pairs.

$$\left\|\widetilde{\mathbf{X}}'_i-\mathbf{X}'_i\right\|_2\leq\varepsilon. \quad (19)$$

In the formula, $\mathbf{X}'_i$ represents the point $(x'_i,x'_j,1)^T$ that matches $\mathbf{X}_i$, and ε represents the residual threshold.

(3) If $M>M_b$, the current homography matrix $\mathbf{H}$ is the optimal solution; that is, $\mathbf{H}_b=\mathbf{H}$. If $\mathbf{M}=\mathbf{M}_b$, compare the cost functions of $\mathbf{H}$ and $\mathbf{H}_b$ calculated by formula (19), and the homography matrix with the smallest cost function is $\mathbf{H}_b$.

After the above steps are iterated N times, $\mathbf{H}_b$ is stored in the set $\mathbb{Q}$, and the subset $\overline{C}_k$ composed of the interior point pairs determined by $\mathbf{H}_b$ in the cluster pair C_k is stored in the set $\mathbb{C}$. Then, the algorithm continues to repeat the above iterative steps for cluster pair C_{k+1} until $k=s$, $\mathbb{Q}=\{\mathbf{H}_1,\mathbf{H}_2,\ldots,\mathbf{H}_s\}$, and $\mathbb{C}=\{\overline{\mathbf{C}}_1,\overline{\mathbf{C}}_2,\ldots\overline{\mathbf{C}}_s\}$ are obtained. The symbol T_m is used to represent the threshold logarithm of the interior points in the cluster pair. If $\mathrm{card}(\overline{C}_k)\geq T_m$, then $\overline{C}_k$ is considered to be a set of matching point pairs; otherwise, $\overline{C}_k$ is discarded. The set of interior point cluster pairs filtered by the threshold $\overline{C}_k$ is denoted as $\mathbb{C}'$, and the set of corresponding homography matrix is denoted as $\mathbb{Q}'$.

Generally, after filtering, there are still a certain number of similar matching point pairs in the text image, and the text image is considered to be a tampered text image. For the algorithm in this article, if $\mathrm{card}(\mathbb{C}')>0$, it means that there are still matching point pairs in the text image after filtering. It means that the text image has been copy-paste-tampered; otherwise, the text image is considered real.

4. Novel Multiple Roles Analysis Model Based on Deep Learning Text Classification

This article combines the deep learning text classification method to construct a novel multiple roles analysis model and verifies the reliability of this method by analyzing the novel multicharacter. The specific ideas of deep learning are shown in Figure 2:

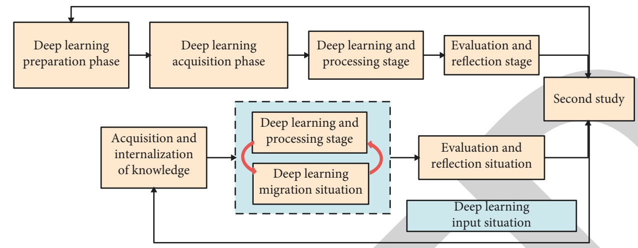


Figure 2: The specific ideas of deep learning.

In natural language processing, the preprocessing of corpus data is also crucial. Automatically extracting language information from the text can overcome the difficulty of acquiring language knowledge. The preprocessing of literary works includes basic tasks such as word segmentation, part-of-speech tagging, named entity recognition, syntactic analysis, name clustering, and reference resolution. The preprocessing process of the novel is shown in Figure 3.

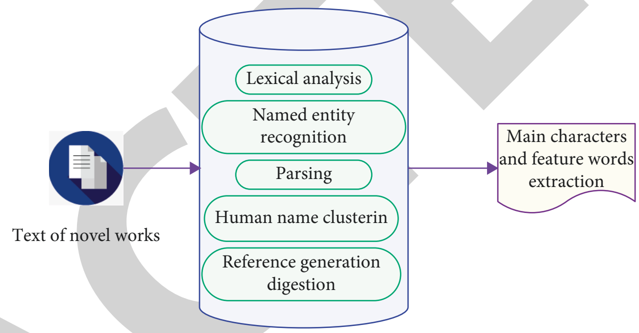


Figure 3: The preprocessing process of the novel text.

This article uses the context based on the dependency relationship as the feature word to train the vector of the novel character. The process is shown in Figure 4.

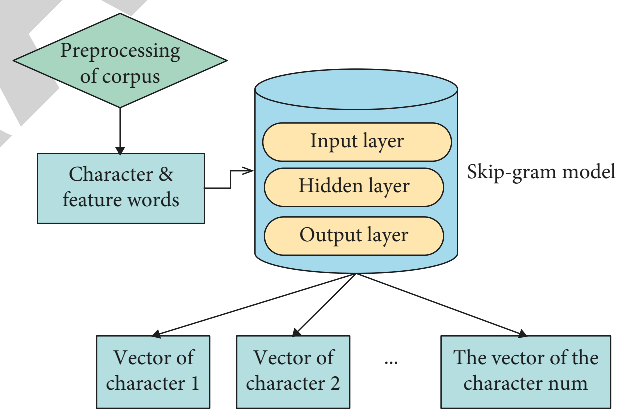


Figure 4: The process of training the character vector.

In the data preprocessing, the results of the syntactic analysis of all texts are counted. It is found that the characteristic words that have a syntactic relationship with the characters in the novel are mainly composed of five types of sentence components, namely, poss (all forms), dobj (direct object), nsubj (noun subject), amod (adjective), and nsubjpass (passive noun subject). Using the parts of speech of these five types of characteristic words, we divide them into four groups and then merge these five types of characteristic words into a whole, resulting in a combined form of five groups of characteristic words. The first group uses the overall combination of five types of feature words to represent the characters in the novel, which is the character vector representation method proposed in this article, and the character vector is denoted as c_pdnan_vec. The second group is to use the words nsubj and nsubjpass to represent the characters in the novel, and the character vector is denoted as c_nn_vec. The third group is to use only words containing dobj to represent the characters in the novel, and the character vector is denoted as c_d_vec. The fourth group is to use only words containing poss to represent the characters in the novel, and the character vector is denoted as c_p_vec. The fifth group is to use only words containing amod to represent the characters in the novel, and the character vector is denoted as c_a_vec. The specific representation method of these five groups of person vectors is shown in Figure 5.

The fast and accurate classification model of long text based on FastText is shown in Figure 6. In the input part of the statistics module, in addition to the prediction result of the sentence block, the weight W of the sentence block needs to be added. Among them, the definition of w is $W = \{wi| 1 \le i \le \text{len (blocks)}\}$, and Wi is the weight of the sentence block whose sequence number is i. After that, in the statistics module, the scores of each category are counted according to the weight, and the score Scoreci of each category and the total score of all categories Scoretotal are obtained. After that, according to the size of Scoreci, the proportion P of the Scoretotal occupied by the category C and Scorec with the highest score and the number index of the key sentence block predicted to be category C are obtained. Finally, the model outputs C, P, and index. The specific algorithm flowchart of the fast and accurate classification model of long text based on FastText is shown in Figure 7.

After constructing the above model, the model is verified, the model is verified in real time, and a large number of novels are obtained through the network, and the model in this article is used for text classification. On the basis of text classification, this article analyzes the multiple roles of novels and builds the model of this article through a simulation platform. The statistical text classification results are shown in Table 1 and Figure 8.

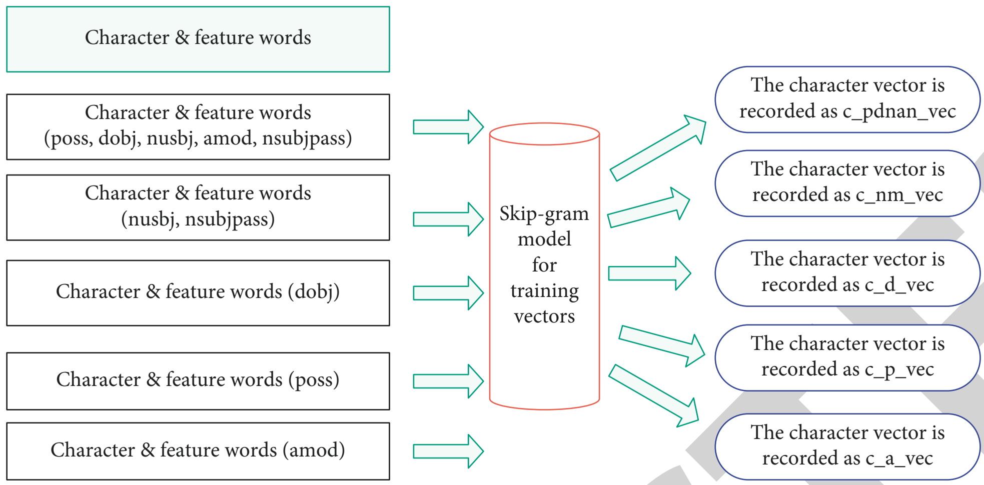


Figure 5: Schematic diagram of training character vectors of different sentence components.

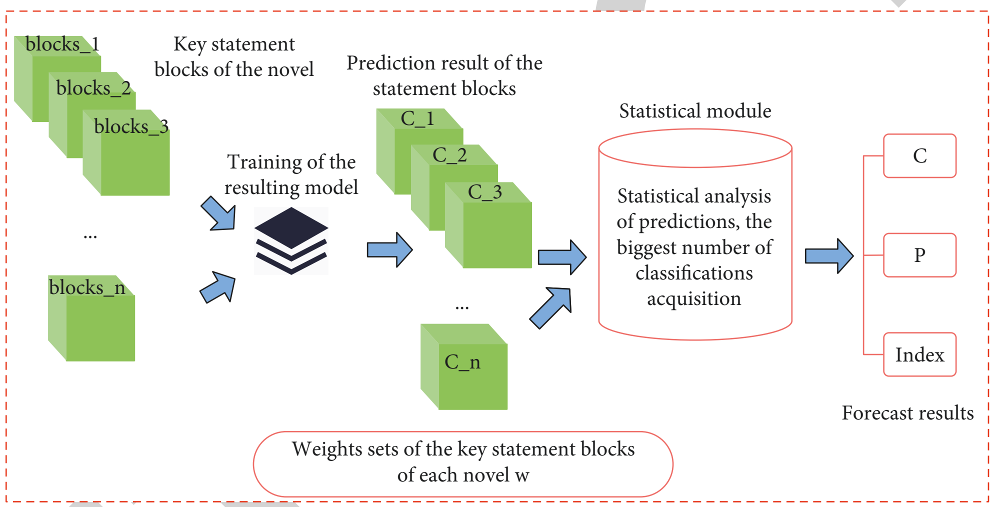


Figure 6: Algorithm flowchart of the fast and accurate classification model.

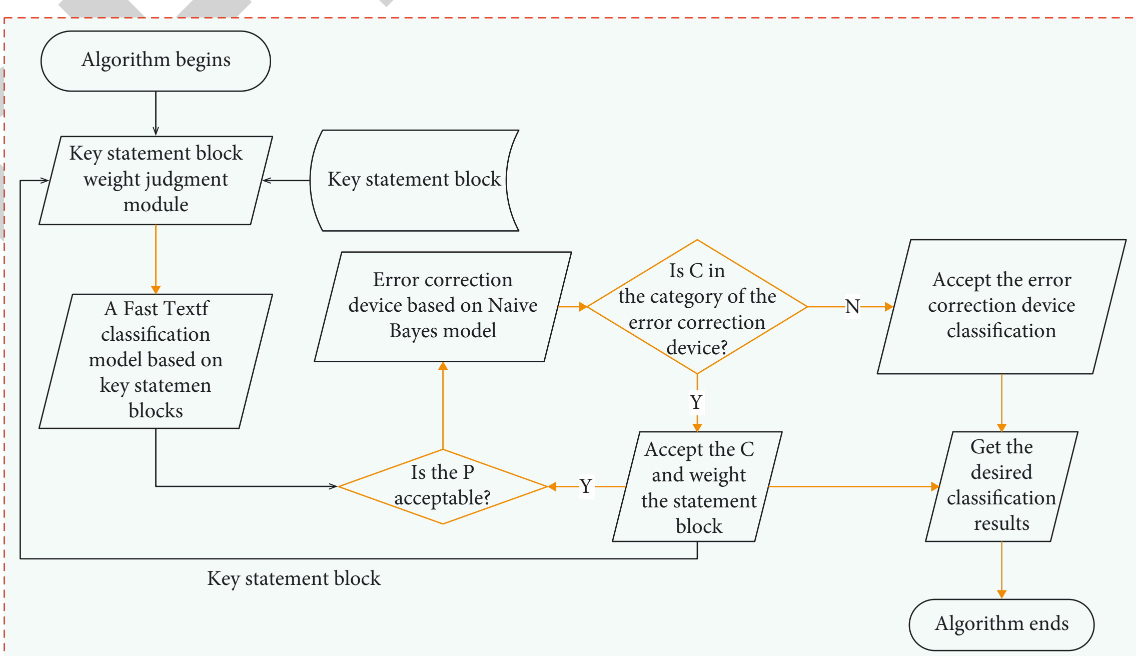


Figure 7: Algorithm flowchart of the fast and accurate classification model.

Table 1: The classification effect of the novel text classification model based on deep learning.

Number	Text categorization	Number	Text categorization
1	89.84	19	84.97
2	89.62	20	91.60
3	85.89	21	91.35
4	87.58	22	91.30
5	84.47	23	85.49
6	91.15	24	85.90
7	86.58	25	86.95
8	86.26	26	84.51
9	90.92	27	84.00
10	84.92	28	91.83
11	87.20	29	88.07
12	91.73	30	84.23
13	84.51	31	86.93
14	85.33	32	87.03
15	85.60	33	87.65
16	85.91	34	86.34
17	86.10	35	85.28
18	88.52	36	91.62

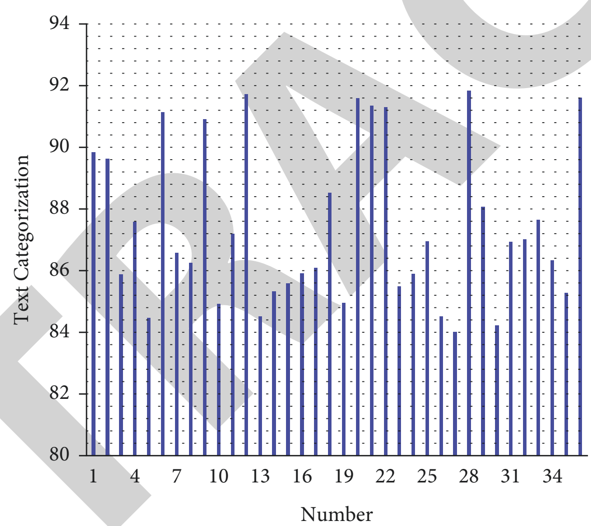


Figure 8: Data statistics of the novel text classification experiment.

Table 2: Statistical table of the effect of novel multirole analysis based on deep learning text classification model.

Number	Role analysis	Number	Text categorization
1	84.16	19	71.95
2	74.33	20	77.35
3	74.67	21	76.33
4	76.90	22	73.97
5	76.91	23	72.97
6	77.20	24	74.86
7	71.10	25	82.19
8	80.65	26	80.66
9	78.36	27	84.03
10	80.05	28	79.45
11	85.16	29	73.12
12	81.58	30	79.74
13	79.18	31	84.19
14	78.48	32	78.45
15	73.75	33	79.81
16	84.58	34	72.44
17	85.72	35	82.30
18	80.68	36	83.64

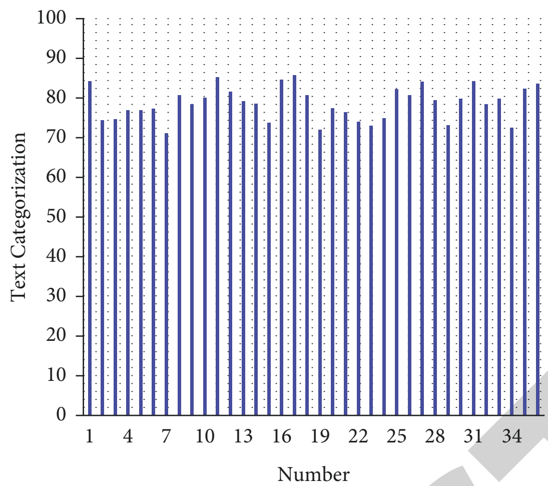


Figure 9: Experimental data of novel multirole analysis based on deep learning text classification model.

Table 3: Comparison of the effects of intelligent text feature analysis models.

Number	The method of this article	The method of [8]	Number	The method of this article	The method of [8]	Number	The method of this article	The method of [8]
1	85.31	75.06	19	80.30	76.01	37	82.34	77.00
2	82.89	69.23	20	77.46	72.12	38	72.61	68.39
3	83.49	71.46	21	72.80	71.03	39	71.40	76.62
4	80.33	70.62	22	78.85	69.65	40	83.05	66.02
5	78.61	73.32	23	74.46	63.86	41	78.38	65.69
6	81.20	65.01	24	84.04	72.09	42	84.70	62.47
7	76.68	66.04	25	79.95	65.10	43	84.33	65.25
8	71.13	72.12	26	79.88	62.23	44	81.20	76.37
9	80.41	76.71	27	85.17	66.79	45	80.60	72.33
10	79.73	64.12	28	77.45	65.10	46	76.16	69.03
11	79.56	65.11	29	75.01	67.41	47	84.05	74.19
12	80.47	76.68	30	81.84	63.21	48	85.27	62.04
13	83.01	73.46	31	74.76	68.95	49	72.01	64.39
14	81.16	75.46	32	73.87	74.49	50	81.65	68.24
15	83.37	63.01	33	85.31	70.74	51	81.54	65.21
16	82.69	72.44	34	80.48	67.03	52	81.91	63.36
17	74.43	75.16	35	75.73	71.99	53	77.37	62.33
18	82.60	66.55	36	78.81	71.41	54	84.52	66.50

It can be seen from the above research that the effect of the novel text classification method based on deep learning proposed in this article is very good, and then the statistics of the effect of the text classification model based on deep learning in the multirole analysis of novels are carried out, as shown in Table 2 and Figure 9.

The intelligent text analysis model proposed in this article is compared with the method proposed in the literature [8], and the effect of intelligent text analysis is calculated. The statistical results are shown in Table 3.

Through the above research, it can be seen that the deep learning text classification model for multiple roles in novels proposed in this article has good effects on role analysis and text classification and meets the needs of intelligent analysis of novels.

5. Conclusion

The research of systemic functional linguistics adopts a quantitative method to analyze the personality of characters in a certain vocabulary grammar framework, word analysis theory, or relatively stable vocabulary. Qualitative analysis of characters in novels is the mainstream of current psychological analysis of characters in novels, and the psychological analysis of characters in novels generally depends on the subjective experience and literary quality of researchers in the research process. At present, deep learning technology has been widely used in many tasks of natural language processing, but the application of text in the field of literature is still less. This article studies the basic problem of fictional character modeling, which is a basic problem of computa-

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9790486, 1 page
https://doi.org/10.1155/2023/9790486

WILEY | Hindawi

Retraction

Retracted: Analysis of Efficiency of Human Resource Management Evaluation Model Based on SOM Neural Network

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and name external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Liu, X. Wang, X. Du, and P. Gu, "Analysis of Efficiency of Human Resource Management Evaluation Model Based on SOM Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 4682868, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4682868, 12 pages
https://doi.org/10.1155/2022/4682868

Research Article

Analysis of Efficiency of Human Resource Management Evaluation Model Based on SOM Neural Network

Xu Liu,[1] Xuewen Wang ,[2] Xiaowei Du,[3] and Peng Gu[4]

[1]*Office of Academic Affairs, Hebei University of Engineering, Handan, Hebei 056038, China*
[2]*School of Earth and Engineering, Hebei University of Engineering, Handan, Hebei 056038, China*
[3]*Organization Department, Hebei University of Engineering, Handan, Hebei 056038, China*
[4]*Information Technology Department, Bank of Handan, Handan, Hebei 056001, China*

Correspondence should be addressed to Xuewen Wang; wangxuewen@hebeu.edu.cn

Received 6 January 2022; Revised 15 February 2022; Accepted 23 February 2022; Published 23 March 2022

Academic Editor: Chin-Ling Chen

The success of human resource management is directly related to whether enterprises can stand in the fierce market competition. In the actual operation of enterprises, the human resource management faces various risks. Here, we take the human resource management evaluation standard as the research object and establish the human resource management benefit evaluation model through a SOM neural network. The training, studying, and testing processes of the model are performed. Besides, the human resource management evaluation system is built on the basis of this model. The evaluation research on the efficiency of human resource management has a good influence in many aspects. Researchers have tried to decipher the "black box" relationship between human resource management and organizational benefits, and an effective human resource management evaluation tool is obtained. This model involves both theoretical construction and empirical research. It is of great significance to summarize the progress and existing problems of current research and abroad and put forward some suggestions for future research.

1. Introduction

Under the joint economic work of the world, competitions between different companies have become more and more obvious. In the actual business process of the company, the degree of dependence on human resources has gradually increased, leading to the problem of various benefit evaluation systems for human resource management [1]. Evaluating efficiency of human resource management has always been a hot and important topic in the field of human resources management. Human resource management efficiency refers to the proportional relationship between the contribution share of human resource management to the benefit of the corresponding organization and the resources consumed by itself [2]. R.S. Schuler proposed an evaluation method for analyzing the benefit contribution of human resource management to organizations. This process should be carried out from two aspects: the evaluation of human-centered results and the evaluation of organization-centered results [3]. Since then, researchers have tried to build a more effective efficiency of human resource management evaluation system. However, there is still no recognized successful efficiency of human resource management evaluation tool. Because of the different perspectives of researchers on the objects of concern and research methods, there are many views in this field. Human resource planning is the process of forecasting the supply and demand of human resources and balancing the supply and demand of employees. It is the premise for the smooth operation of enterprises and the basic guarantee for enterprises to obtain competitiveness. The two are complementary to each other in the development of enterprises [4].

This paper reviews the research on efficiency of human resource management evaluation at home and abroad, analyzes the achievements and shortcomings of related research, and discusses the direction of improvement, in order to provide reference for future research and promote the development of human resource management theory [5].

From the perspective of research methods, the current research on efficiency of human resource management evaluation can be roughly divided into two categories: one is theoretical exploration from a qualitative perspective, and various theoretical models are proposed from different perspectives, which are mainly constructed from single linear, nonlinear, and systematic perspectives. The other type is empirical research from a quantitative perspective according to the relevant theoretical framework [6–9].

Self-organizing map (SOM) neural network generates a low-dimensional and discrete map by learning the data in the input space, which can also be regarded as a dimension reduction algorithm to some extent. SOM is an unsupervised artificial neural network. Different from the normal neural network training based on the reverse transfer of loss function, SOM uses a competitive learning strategy to optimize the network step by step depending on the competition between neurons [10]. It can quickly learn the laws that exist in a set of data, classify them in discrete time, and map high-dimensional data to low-dimensional space one by one. This makes the internal similarity of input data show the feature mapping of spatial neighbors. Through this method, a one-dimensional or two-dimensional discrete graph can be mapped, and its topological structure remains unchanged. SOM is a system unit composed of input and output parts. The first part contains one-dimensional spatial elements of k nodes, and the other part is two-dimensional spatial elements. Node matrix is composed of $M = m_2$ nodes. Yuan is connected by a certain weight [11]. It can be seen that what Dyer, Macduffie, and Becker have in common is that they all think that the influence of human resource management practices on organizational benefits is intuitive and a simple linear causality model, and the latter variable in the model is only influenced by the former variable. However, in reality, the actual situation of enterprises is much more complicated. There are many variables that are influenced by human resource management practices and ultimately affect the performance of enterprises, and these models do not consider the influence of other variables. Therefore, this simple and linear mold is not perfect [12]. Ferris put forward a model of the relationship between human resource management and organizational benefits under the social background. The theory extended the antecedents and added more intermediate influencing variables, which made the process of human resource management practice affecting organizational benefits more complicated. This model starts with organizational culture, which is crucial to the formation and type of human resource management (HRM) system and its influence in the implementation process, and HRM system contributes to organizational benefits through flexibility, employee behavior, and organizational prestige. HRM system can shape the flexibility of the organization, thus improving the efficiency of the organization [13]. SOM neural network is an intelligent neural network, a learning tool without external supervision. Human resources are such resources. When it penetrates deeply into the operation system of the organization, not only can it create value and increase the strength of the enterprise, but also this competitive advantage is difficult to imitate. Therefore, through the establishment of such a model, it is possible to intuitively increase the competitiveness of enterprises in various industries through technical means. When the human sensory organs are stimulated, the brain releases specific neurons to make the human body start to excite. As shown in Figure 1, the network is similar to this mechanism, approaching the output state. The corresponding network structure is applied in this study [14]. Although more and more enterprises realize the importance of human resource management, how to measure the performance of human resource management department has always been a big problem that plagues enterprises. This paper is to establish a model of human resource management benefit evaluation on the basis of SOM neural network, so as to obtain a series of human resource management information.

2. Construction of Evaluation Model

The schematic diagram of the human resource management efficiency evaluation model of SOM neural network is shown in Figure 2.

The following steps are essential to establish the efficiency evaluation model of human resource management based on SOM neural network:

(1) Analyze the data to determine the set of risk elements.

The company can choose suitable human resource benefit factors as the evaluation criteria, as Zheng used the formula in the algorithm of his article [15]:

$$N_{mn} = e^{-(Wm - W_n)^2/2\delta^2}. \quad (1)$$

In the above formula, N_{mn} is the evaluation standard function, and δ is a speed constant in the training process.

(2) Select the input information points of SOM neural network.

When using this model, the validity and accuracy of its predictions will depend on the selected coefficients. When the human resource management benefit is the object, the set of elements determined in the previous period is used to automatically transform the range into a closed interval from 0 to 1. Using normalized calculation method to deal with may cause ambiguity in the result, as Chen used the normalization function formula in his article [16]:

$$P_{mn} = \frac{W_{mn} - W_m}{W_{mn} - W_n}. \quad (2)$$

In the above formula, $W_m = \text{mix } (W_{mn})$, $W_n = \max (W_{mn})$, and $P_{mn} \in [0,1]$ is the result of normalization function.

(3) Select benefits of SOM neural network.

In this paper, the accuracy of its prediction efficiency is mainly determined by the correctness of the information point selection. There is a direct ratio between the two. If a few information points are selected, the information output by SOM system will be further reduced and the accuracy of the network information will also be reduced.

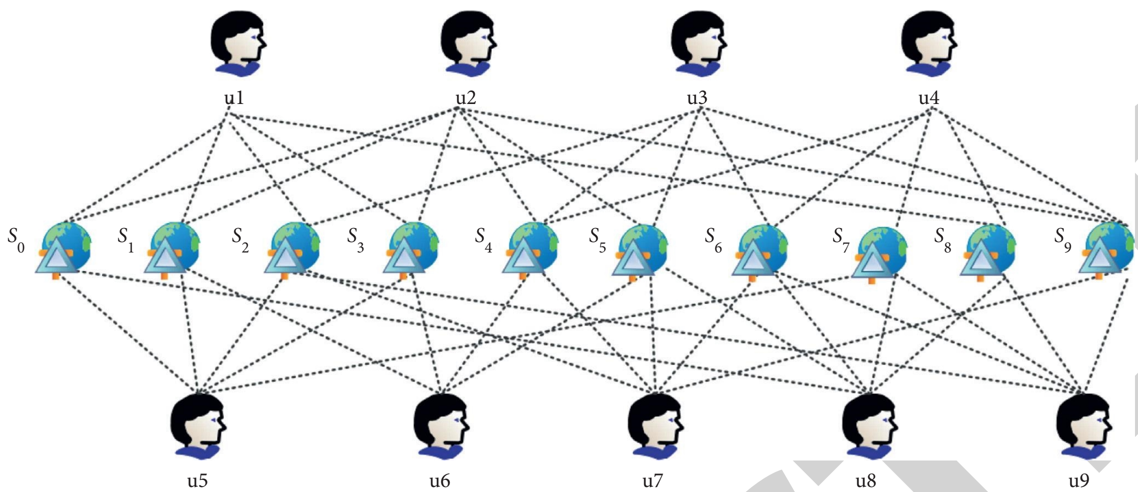


Figure 1: Relationship between human resource management and organizational benefits.

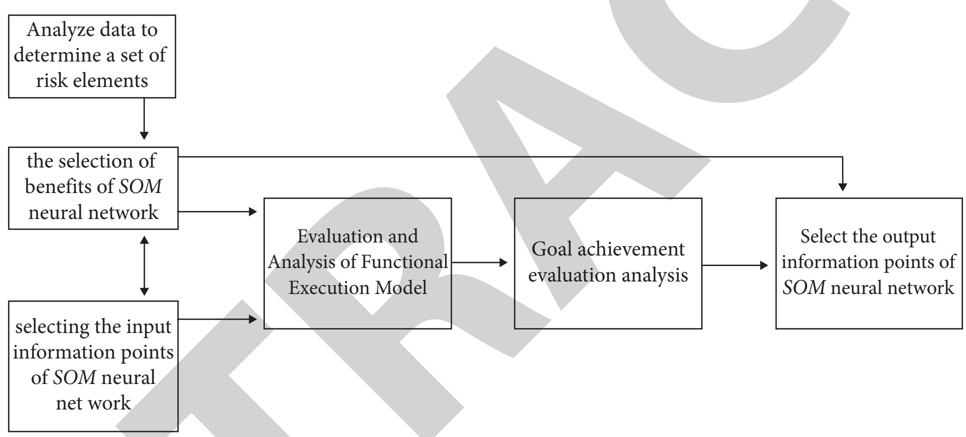


Figure 2: Schematic diagram of the algorithm flow.

Equation (3) represents the criterion for the number of information points obtained in this model.

$$Q = \frac{\sqrt{\sum_{n=1}^{m} (W_m - W_n)}}{P - 1}. \tag{3}$$

In the above equation, Q is the best number of hidden information points, and P is a random constant. The best training time can be obtained by calculation.

(4) Select the output information points of SOM neural network.

The output value of the system can be directly used to evaluate the results of the model. If the evaluation models have different levels, these can be classified into five different levels, namely, Grade I (safest), Grade II (safe), Grade III (basically safe), Grade IV (dangerous), and Grade V (most dangerous). It can be shown by the following formula [17]:

$$H = \sum_{t=1} (j_i - j_t) L(j_i, j_t). \tag{4}$$

In the above formula, H represents the number of information points selected, and j represents the size of the evaluation element.

In the experiment, in order to evaluate the effectiveness of the evaluation method of SOMQP proposed in this paper, we use *WS-Dream* dataset [18], in which two matrices, respectively, contain the response time (RT) and throughput (TP), in which 83.7% of the response time is between [0,1]. Figure 3 shows a clustering process diagram. In fact, many users only select a small number of service systems, so the generated user-service matrix is an augmented matrix. Therefore, we will randomly remove some data and use the rest as a test set. In *WS-Dream* dataset, the proportions of training set are 0.05, 0.1, 0.15, 0.2, and 0.25.

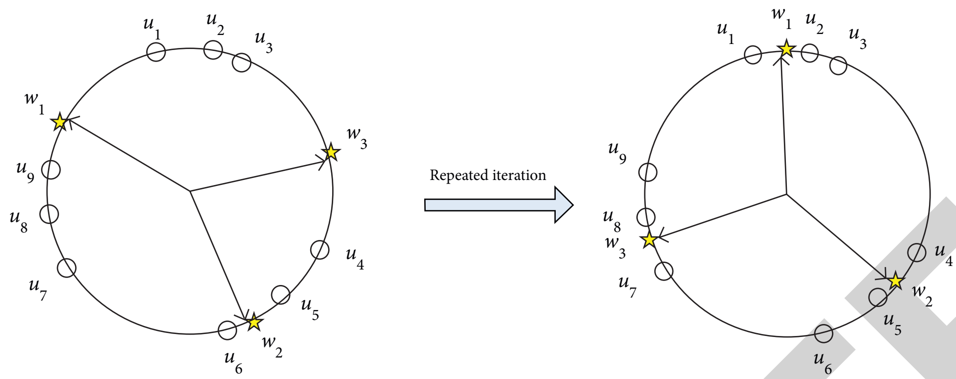


Figure 3: Clustering process.

3. Advantages of SOM Neural Network Evaluation Model over Traditional Evaluation Model

3.1. Traditional Evaluation Model. In the past evaluation practice, the fuzzy comprehensive evaluation method was often used after determining the evaluation index set. The SOM clustering model is a very important category of artificial neural network, which plays an important role in data mining. This model uses a competitive learning strategy to optimize the network to achieve the optimization of the algorithm structure. The next step is to determine the dimensionless characteristic value of each index and the weight of each index relative to the previous index. General expert scoring method is used to determine the weight and then select the comprehensive evaluation model. The general weighted average model was selected, and the corresponding evaluation index was set. Finally, the evaluation result was obtained by summation [19].

However, this method is not suitable for efficiency of human resource management's evaluation. Because efficiency of human resource management is the result of many factors, there is mutual influence among various influencing factors. To truly reflect the efficiency of human resource management, an explicit expression function about the relationship between various influencing factors and results was built with the typical nonlinear system. The SOM neural network, which can effectively solve the dynamic and nonlinear problems in the evaluation process, was unshed in this study as can be seen from Figure 4.

In addition, there are some disadvantages in using expert scoring method to determine the weight of each index. First, the subjective component is big. The rationality of weight determination may be interfered in by external factors. Another disadvantage is that once the weight is determined, it is hard to change. This is not in line with the current changing environment [20].

3.2. Evaluation Method Based on SOM Neural Network. In the mathematical theory of artificial neural networks, the universal approximation theorem points out the ability of artificial neural networks to approximate arbitrary functions. Usually, the neural network referred to by this theorem is a feedforward neural network, and the approximated objective function is usually a continuous function whose input and output are both in Euclidean space. However, there are also studies extending this theorem to other types of neural networks, such as convolutional neural networks, radial basis function networks, or other special neural networks. If only the lowest index, that is, the basic evaluation index, is considered as the input and the middle index is not considered, the whole evaluation system is regarded as a black box, and finally an output, that is, the benefit degree, is obtained. In this way, the whole evaluation system can be regarded as a black box of complex nonlinear functions, which actually has something in common with the mesh evaluation index system, and also conforms to the characteristics that human resource management is a large system with complex mechanism [21].

For example, in the tree-type evaluation index system, we regard "safety accident rate" as a basic index under the improvement index of work and life quality in the "goal achievement degree" subsystem, because the safety of work is one aspect of work quality. But, at the same time, the reduction of safety accident rate may be the result of training or the function of motivation [22]. Figure 5 shows the clustering of input vector neurons with different benefits. Therefore, "safety accident rate" is not only related to the improvement index of work and life quality but also related to the reserved index of human resources.

Based on this black box idea, this paper introduces neural network as the realization model of efficiency of human resource management evaluation system [23]. The system is a multielement network formed by a large number of interconnected neuron units containing simple information. Basically, it can simulate the structure and reaction of human brain nerve. The neural network with highly nonlinear characteristics was built based on an ultra-large-scale self-processing information system. Theoretically, the network space of two levels can be infinitely close to any nonlinear function, so the feedforward neural network can be regarded as an unprecedented powerful learning system. The introduction of neural network provides a new idea for determining the weights of efficiency of human resource management evaluation system, because neural network has many excellent qualities such as adaptive self-organization and being good at making decisions from the approximate uncertain or even contradictory knowledge environment can avoid artificially selecting weights and calculating correlation coefficients.

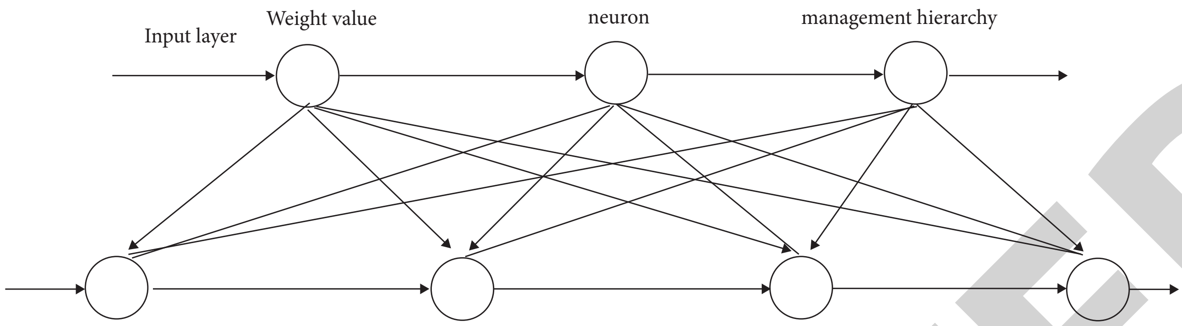


Figure 4: Efficiency of human resource management model based on SOM neural network.

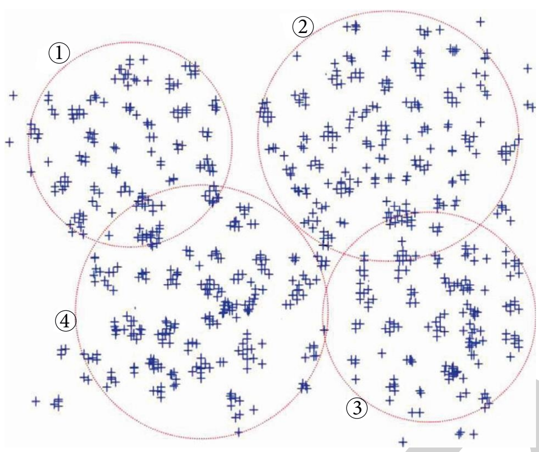


Figure 5: Clustering of input vector neurons with different benefits.

In the neural network, there are many kinds of algorithms. Among them, the SOM network is one of the most widely used. Generally speaking, it consists of two parts: the input layer and the output layer, and the two are completely interconnected. However, the units on the same floor are not connected [24]. The learning process of neural network includes two propagation forms: forward propagation and backward propagation. The former inputs the samples from the input layer after being processed by the hidden layer and then transmitted to the output layer. If the output format does not meet the requirements, it will become another spread. In this process, the error signal propagates from the output layer to the input layer, which will change the connectivity weight of each layer and the bias value of each layer of neurons, so that the error signal will continue to decrease in this process. Figure 6 shows management cluster identification based on self-organizing mapping. After repeated changes, when the error range meets the requirements, the network training process is also over.

3.3. Scope of Application of Two Methods in Internal Control Evaluation. Fuzzy comprehensive evaluation method can be used for comprehensive evaluation of subjective factors as well as objective factors. In the process of internal control evaluation, there are a lot of fuzzy phenomena of evaluation objects, especially in comprehensive evaluation with many subjective factors. Because subjective factors are very fuzzy, using fuzzy comprehensive evaluation can give full play to the advantages of fuzzy methods. However, on the other hand, due to the flexibility of artificial weighting, the weights of evaluation factors are different because of the different focus of evaluators [25]. Due to the subjectivity of human beings, there may be deviations between the determination of weights and objective reality, which may affect the accuracy of evaluation results. It is necessary to form a composite evaluation system with the help of other methods such as SOM neural network evaluation method to correct the defects of single method evaluation and improve the accuracy of internal control evaluation results.

SOM neural network evaluation method has the characteristics of being high-speed self-learning and self-adaptive, fault tolerance, flexibility, and so on, and the system for evaluating information imperfection has more advantages. Figure 7 shows the changes of the two economic benefits with time. When the analysis object of internal control evaluation is fuzzy, incomplete, and uncertain, the sample data can be used for sufficient training and testing, and the evaluation results can be obtained by effective training. However, on the other hand, neural networks often converge slowly, and training takes a lot of time. Because of the complex network structure and algorithm, it is difficult to understand and master, so it has higher requirements on the technical level. Moreover, when there are too many influencing factors and levels of internal control evaluation objects, with the increase of training times, the calculation and storage capacity will increase, and there may be overfitting, so that accurate evaluation results and predicted values cannot be obtained.

It has the advantage of evaluating multilevel complex problems. Fuzzy comprehensive evaluation analyzes complex objects from the perspective of hierarchy. The more complex and hierarchical the structure of the evaluated object is, the better the effect of multilevel fuzzy comprehensive evaluation will be. The result of fuzzy comprehensive evaluation is represented by a fuzzy set, which accurately depicts the fuzzy state of the thing itself. After further processing and giving an appropriate score, an equal score vector can be calculated, which provides a quantitative method for qualitative problems [26].

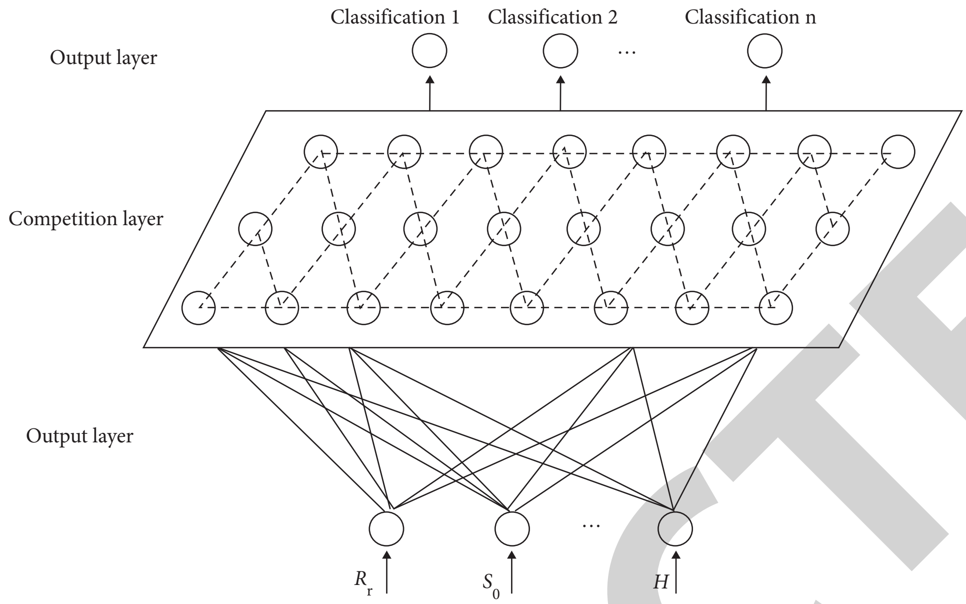


Figure 6: Management cluster identification based on SOM.

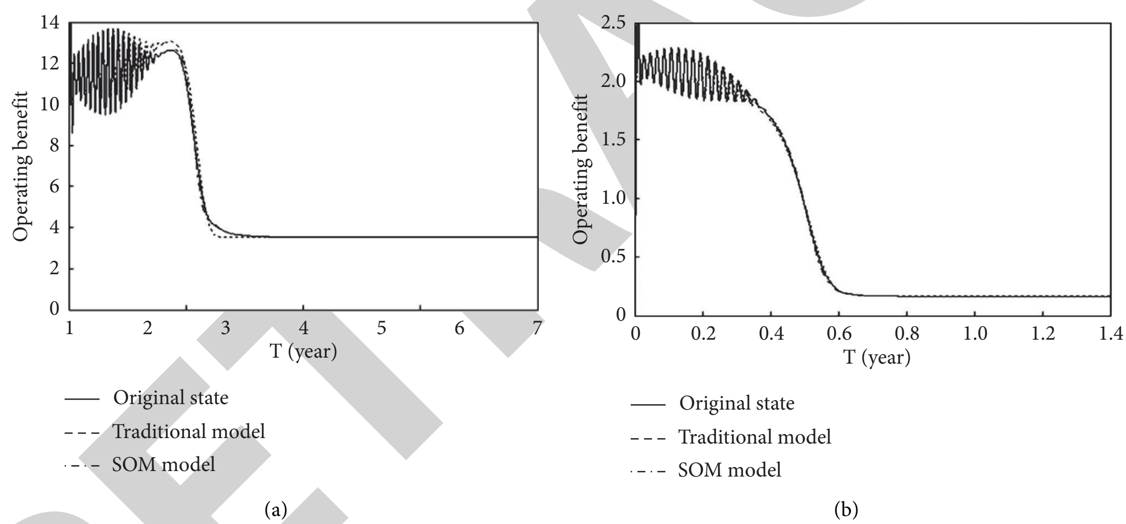


Figure 7: Changes of two kinds of economic benefits with time.

With strong fault tolerance, it can fully approach any complex nonlinear curve, and samples are not required to be independent or obey normal distribution, and all quantitative or qualitative information is distributed and stored in each neuron in the network. It has storage function and self-learning function and can learn and adapt to unknown or uncertain systems.

4. SOM Clustering and Top-k Selection Mechanism

Let the input layer $x = (x_1, x_2, ..., x_n)$ be an N-dimensional vector, let the output layer be a two-dimensional network with m nodes, and let w_{ij} be the weight between the *i-th* input neuron node and the *j-th* output neuron node. The training process of this algorithm was realized under the neural network environment [27]. Figure 8 shows the comparison of experimental results of CPI prediction.

(1) Initialize the connection weight, learning efficiency, and neighborhood, and w_{ij} selects random values between [0,1], and all values are different from each other.

(2) Normalize the sample and the connection vector. After normalization, calculate the input vector in the distance *d(x,w)* from the output node:

$$d(x,w) = \sqrt{\sum_{i=}^{n} \left(x_i - w_{ij}\right)^2}. \tag{5}$$

(3) Select the smallest *d(x,w)* from the distance nodes calculated above as the best matching node; that is, neuron i is the winning neuron.

(4) Adjust the update formula of weight vector:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)h(t)\left[x(t) - w_{ij}(t)\right]. \tag{6}$$

In the above formula, $\alpha(t)$ is the learning efficiency, $0< \alpha(t) < 1$, and it decreases with time t; $h(t)$ is a function of the topological distance between the ith neuron and the winning neuron j in the neighborhood.

(5) Repeat the above steps until the learning efficiency $\alpha(t)$ is less than α min after learning all samples.

According to the above-mentioned clustering principle, for different systems that call the same service, if the QoS values in their model systems are the same or similar, then when users are clustered, there will be a greater probability that they will be clustered in the same cluster.

Before clustering, initialize the user relationship matrix M_u so that the initial values of all elements are 0. Aiming at the user set U_s calling service S, the SOM algorithm is used to cluster, and the weights w_1, w_2, and w_3 of neurons are initialized [28]. Firstly, the Euclidean distance between the first data (the score of the first user u_1 on service S) and three neurons is calculated, and neuron 1 with the smallest distance wins, and then weight w_1 of the winning neuron 1 is changed according to formula (6). Continue to calculate the distance between the next data and three neurons, and so on; each time only one neuron wins and changes the corresponding weight. Until the learning efficiency is less than the threshold, weights w_1 , w_2 , and w_3 of neurons tend to the clustering center, and the iteration stops [29, 30]. One is to focus on the HR value chain model, where HR practices affect employee output, that is, the relationship between attitudes and behaviors. The other is to use a mechanistic reductionist approach to measure the relationship between human resource management and organizational effectiveness using a cumulative stacking method. Then, according to the results of SOM clustering, the elements in the user relationship matrix are updated. The updating rules are assuming that user I and user J are clustered in a cluster in service S; then $u_{ij} = u_{ij} + 1$, and, at the same time, $u_{ji} = u_{ji} + 1$. Otherwise, it remains unchanged.

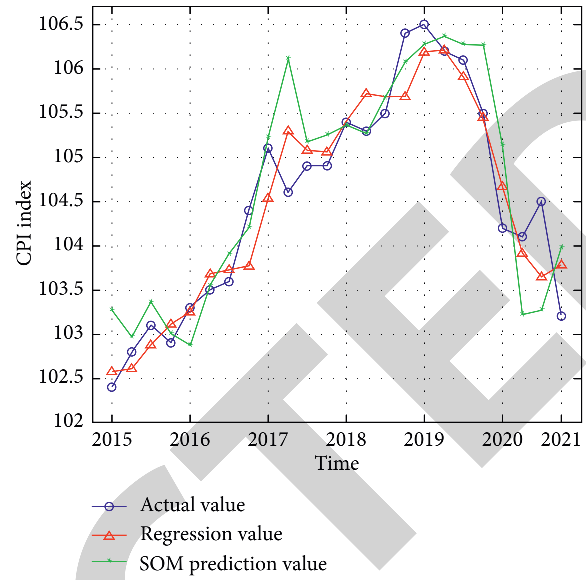


Figure 8: Comparison of experimental results of CPI prediction.

$$R_{u,i}(u) = \frac{\sum_{t=1}^{k}\left(r_{a(t)} \times u_{a(t)}\right)}{N_u}. \tag{7}$$

In the above formula, $u_{a(t)}$ represents the number of times of being clustered into the same cluster.

$$R_{u,i}(u) = \frac{\sum_{t=1}^{k}\left(r_{b(t)} \times u_{b(t)}\right)}{N_s}. \tag{8}$$

In the above formula, N_s represents the sum of the times that service I and all of them are clustered in the same cluster.

$$R_{u,i} = \lambda \times R_{u,i}(u) + (1-\lambda) \times R_{u,i}(i). \tag{9}$$

Parameter λ indicates the proportion by which the predicted value depends on similar users and similar services.

In this paper, a new Top-k selection mechanism is proposed. According to the clustering strategy, if two evaluation models are clustered in the same cluster more times, it is shown that the two models have the same or similar evaluation on the same efficiency of human resource management, and when they call the same benefit, the service evaluation between them is more referential. During clustering, u_{ij} is used to record the times that users u_i and u_j are clustered in the same cluster. After obtaining the user relation matrix M_u, the elements of each row are sorted from big to small, and the first k users are selected as similar users of the target user; that is, the similar user set of users u_i is $F(u_i) = \{a \mid u_{ia} \geq u_{ik}, u_{ia} > 0, A \in U, u_{ik} \in M_u\}$. This can reduce the influence of users' evaluation of malicious evaluation on prediction results. Take users of service s_1 as an example; users u_1, u_2, u_5, and u_7 have called service s_1 and given scores of 4.2, 4.5, 4.4, and 1.1, respectively. From the scores, it can be inferred that user u_7 is probably a user of malicious evaluation. When clustering, user u_7 will not cluster with users u_1, u_2, and u_5.

$$\begin{aligned} \mathrm{MAE} &= \frac{\sum_{u,s}|R(u,s)-P(u,s)|}{N}, \\ \mathrm{RMAE} &= \sqrt{\frac{\sum_{u,s}(R(u,s)-P(u,s))^2}{N}}. \end{aligned} \tag{10}$$

In the two formulas, mean absolute error (MAE) and root mean square error (RMSE) are defined.

Similarly, in clustering services, the more times two human resource evaluation models are clustered into the same cluster, the greater the correlation between the two models is and the more referential the user evaluation between them is. Therefore, after the clustering, the element values of each row in matrix M_s are sorted from big to small, and the first k services are selected according to the similarity of the target, that is, the similar service set of S_i.

5. Problems Existing in the Management of Enterprise Personnel Resources

5.1. The Development and Management of Human Resources Are Relatively Backward. The development of enterprises has been influenced by China's economic system for a long time, and the center of its operation and management is mainly reflected in the management of internal funds, material resources, and technology, while the importance of human resources management is relatively neglected. Its management and philosophy are relatively backward, so there is a lack of practical work on employees' career planning in the actual management of human resources and resources. Figure 9 shows the change of actual and predicted CPI values with time. As far as employees in enterprises are concerned, enterprises have neglected the cultivation of their business and the promotion of their abilities, which has made a definite impact on their all-round development. As for the state-owned enterprises, if they lack professional and far-sighted professionals, they will not be able to fully display the actual effectiveness of their internal employees, and the economic benefits of the enterprises will be affected.

5.2. Insufficient Investment in Human Capital. According to the development status of enterprises, most state-owned enterprises have sufficient capital and advanced equipment and technical means. However, due to the influence of their own system, market environment, and other factors, the investment in human capital has been neglected, which leads to the phenomenon of insufficient professionalism and low work enthusiasm of employees, which affects the development of enterprises. Specifically, most state-owned enterprises often use wage increase and other forms to stimulate employees' work enthusiasm and then obtain economic benefits. Figure 10 shows a typical U matrix diagram. This kind of management can effectively stimulate employees' work enthusiasm and improve their job satisfaction in a short time, but this kind of material incentive is not conducive to the improvement of employees' own ability, and it is not applicable in the long-term development of enterprises.

Enterprises need to pay more attention to the training of employees and make good use of other incentives besides material incentives to ensure the realization of employees' self-worth. For example, through regular training, education, and long-term study, employees' acquisition ability can be improved, the advanced nature of employees can be ensured, and their self-worth can be realized; that is, they can get spiritual satisfaction besides material rewards, further improve their work enthusiasm, maximize their practical utility, and provide guarantee for improving the economic benefits and all-round development of enterprises.

5.3. Single Human Resource Management Mode and Centralized Management Authority. Enterprises pay more attention to the development of management, which leads to the problem that the power of human resource managers is too concentrated. On the one hand, the enterprise pays attention to the management but neglects the management and training of other employees in the enterprise. The single management mode of human resources is not conducive to the comprehensive development of state-owned enterprises. Also the state-owned enterprises pay too much attention to the management, and other employees within the enterprise have few opportunities to learn and promote, which greatly inhibits the development of other employees, which has a certain negative impact on other employees and the overall development of the enterprise.

6. Based on SOM Model, the Strategy of Human Resource Management Is Obtained

6.1. Improve the Development and Management Concept of Human Resources in Chinese Enterprises. Figure 11 shows the cluster diagram of 14 parameters and indicators. Knowledge is flexible, and the management of human resources in state-owned enterprises is of great importance to the development of enterprises. State-owned enterprises should keep up with the pace of the times, bring forth the old and the new, create new human resources, manage thinking and thinking, and attach great importance to the self-development of employees, so as to implement the education and training of internal employees and further ensure the cultivation of employees' working attitude and working ability.

6.2. Increase Investment in Human Capital and Strengthen the Construction of Personnel Team. State-owned enterprises should increase the investment of manpower and capital and put more precision, manpower, and financial resources into the education and training of employees. They should attach great importance to the education and training of employees and strengthen the cultivation and construction of professional teams, so as to maximize the value of employees and realize the rapid and sound development of enterprises. First of all, human resource managers in state-owned enterprises should be prepared for the long-term development, pay attention to the long-term development of internal employees, and increase the training of internal employees' ability and professional quality through regular training,

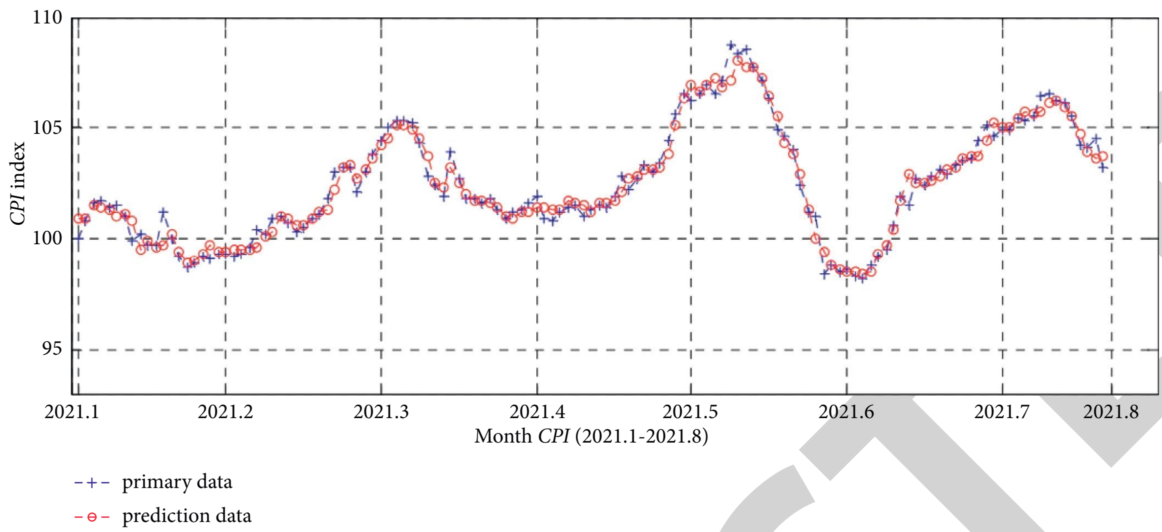


FIGURE 9: Changes of actual and predicted CPI values with time.

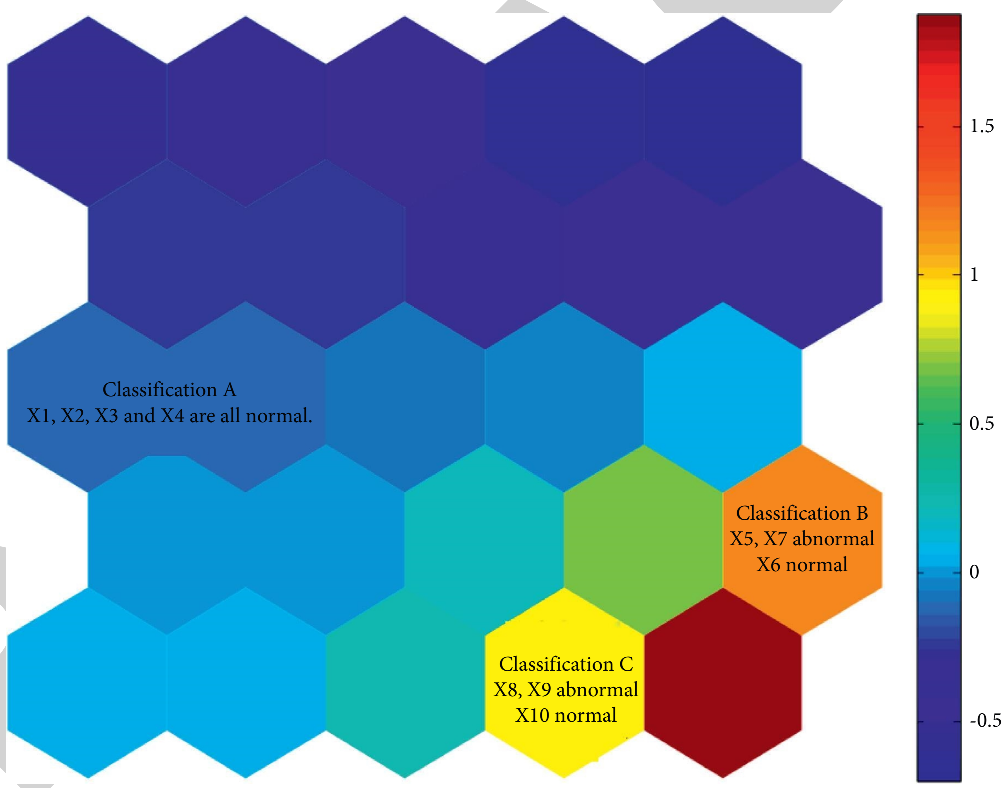


FIGURE 10: Typical U matrix diagram.

educational activities, and going out for further study. Secondly, enterprise human resource managers should also consider their business ability from the perspective of the actual situation of employees, make the best use of their talents, arrange their work positions reasonably, ensure the rational allocation of human resources, improve the efficiency of work, and promote the realization of maximizing the economic benefits of enterprises.

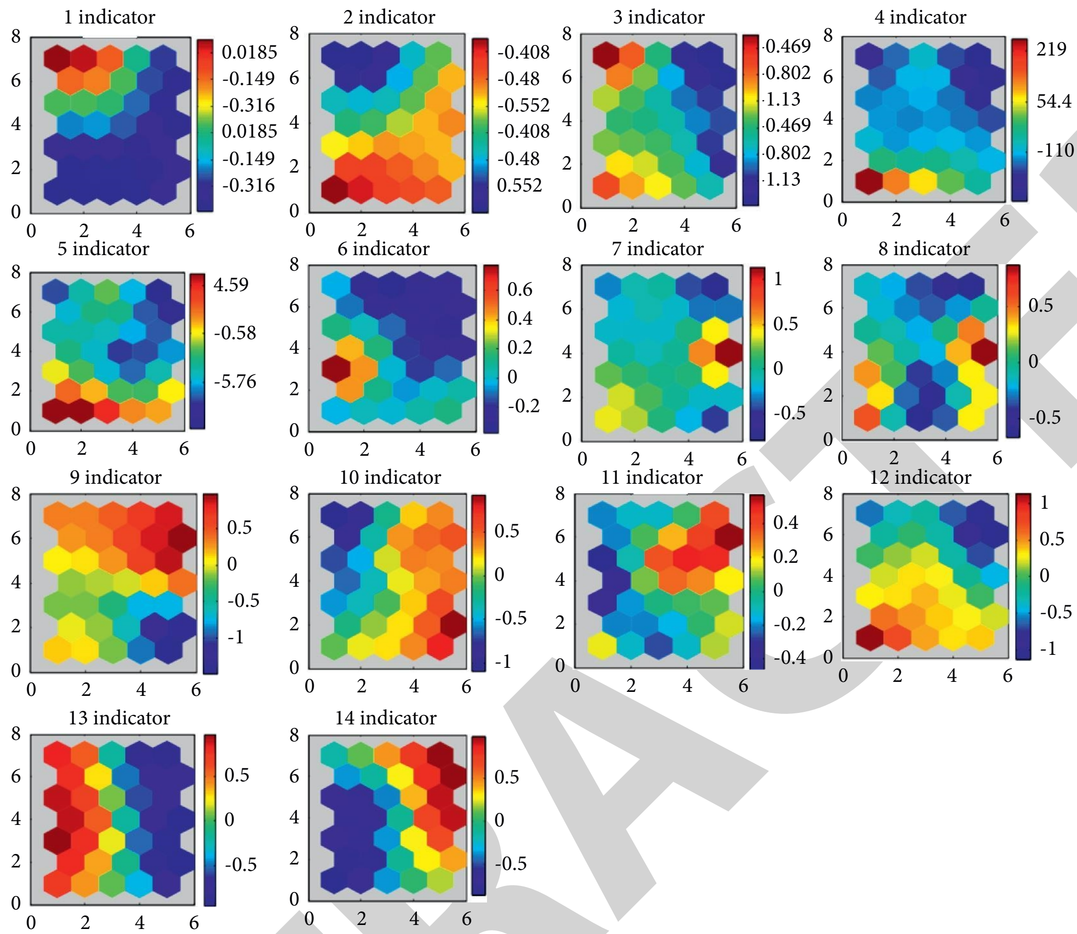


Figure 11: Cluster diagram of 14 parameters and indicators.

6.3. Enrich Management Mode and Decentralize Management Power. Enterprises should change to a single and centralized management mode. Human resource management departments should pay attention to the development of employees' abilities and let employees have room for improvement and promotion, decentralize the traditional centralized management to all levels of organizations and departments within the enterprise, and, at the same time, delegate some of the power to more excellent employees. As shown in Figure 12, the statistical result of Consumer Price Index (CPI) value supports the view. On the one hand, it gives employees room for improvement; on the other hand, it plays a positive role in stimulating employees and improves their enthusiasm for work. For the development of enterprises, the improvement of their enthusiasm for work directly promotes the improvement of work efficiency and further promotes the economic benefits of enterprises.

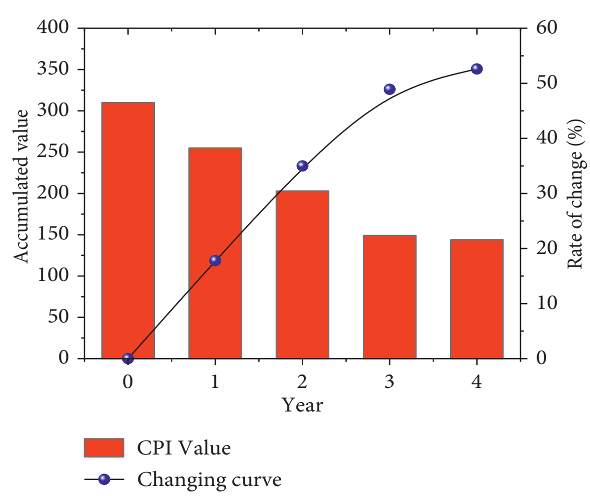


Figure 12: Statistical result of CPI value.

7. Conclusion

In the fierce market competition of modern enterprises, human resource management has gradually become an important research field. In order to understand the evaluation mode of human resource management in time, it has become a problem that managers urgently want to solve. The efficiency of human resource management evaluation model established by this network system can fundamentally ensure that all problems in the management system can be solved. It makes a certain contribution to the progress and stability of the enterprise economy.

In a word, there are different assumptions and methods for evaluating the benefits of human resources management of organization, but there is no recognized successful tool until the end of the day. The evaluation of human resources management, effectiveness, and benefits is either centered on the inner part of the function or centered on the outer part of the function, without combining the inner part and the outer part to evaluate, which leads to the incomplete evaluation index. We believe that, on the premise of fully clarifying the relationship and mechanism between human resource management and organizational benefits and comprehensively considering the "effectiveness" evaluation research of this management, it may be the direction of future research and development to establish a systematic efficiency of this management evaluation system from the internal and external aspects of the organization's functions, using the evaluation method of combining qualitative and quantitative, as well as following the idea of "evaluation-feedback-improvement."

Although the research on this topic has made a definite progress, there are still three main problems. However, a large part of the research is only to analyze the questions that should be paid attention to when measuring the benefits of human resource management. The proposed evaluation index lacks the corresponding support of strength and quantity and has no strong convincing power in its effectiveness. Second, most related research measure the adaptability, implementation, and effectiveness of human resource management in organizations from the three dimensions of coordination, efficiency, and effectiveness, without evaluating the human resource management activities from a systematic angle, as well as paying attention to a certain link of human resource management only, failing to systematically and comprehensively measure the benefits of human resource management in organizations. The dimensions of coordination, efficiency, and effectiveness rarely involve the measurement of various skills of human resource management workers.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper was supported by Hebei Province Higher Education Teaching Reform Research and Practice Project, research on the optimization of teaching quality evaluation system of geology specialty from the perspective of engineering education certification (No. 2021GJJG210).

References

[1] X. Yan, X. Deng, and S. Sun, "Analysis and simulation of the early warning model for human resource management risk based on the BP neural network," *Complexity*, vol. 2022, Article ID 8838468, 11 pages, 2020.
[2] J. Xu, B. Wang, and G. Min, "Research on human resource allocation model based on SOM neural network," *International Journal of Mobile Computing and Multimedia Communications*, vol. 10, no. 1, pp. 65–76, 2019.
[3] R. Gherman, B. Adamescu, I. Brad, and A. M. Dincu, "Efficiency of human resources at national and multinational companies," *Journal of Biotechnology*, vol. 23, no. 1, p. 88, 2016.
[4] B. Jia, S. Liu, and Y. Yang, "Fractal cross-layer service with integration and interaction in internet of things," *International Journal of Distributed Sensor Networks*, vol. 10, no. 3, Article ID 760248, 2014.
[5] Z. S. Lassi, N. B. Musavi, B. Maliqi et al., "Systematic review on human resources for health interventions to improve maternal health outcomes: evidence from low- and middle-income countries," *Human Resources for Health*, vol. 14, no. 1, p. 10, 2016.
[6] X. Li and D. Zhu, "An adaptive SOM neural network method to distributed formation control of a group of AUVs," *IEEE Transactions on Industrial Electronics*, vol. 51 page, 2018.
[7] W. Li, "Marginal utility function based optimal human resource management model," *Systems Engineering —Theory & Practice*, vol. 36, no. 1, pp. 106–112, 2016.
[8] W. Lin, "Human resources management of track and field web course in college physical education," *International Journal of Emerging Technologies in Learning*, vol. 11, no. 4, p. 95, 2016.
[9] L. López-Torres and D. Prior, "Centralized allocation of human resources. An application to public schools," *Computers & Operations Research*, vol. 73, no. C, pp. 104–114, 2016.
[10] L. Marshall and G. Treuren, "Dimensions and determinants of declining employment opportunities for mature aged male practitioners within the human resources profession: occupational change, age and gender," *Research in Nursing & Health*, vol. 18, no. 2, pp. 85–95, 2016.
[11] T. Li, "A demand estimator based on a nested logit model," *Operations Research*, vol. 59, no. 1-2, pp. 107–109, 2019.
[12] J. Wu, Y. Jiang, and J. Zhu, "Human resource allocation combined with team formation," in *Proceedings of the International Conference on Computational Intelligence and Applications*, pp. 67–71, Jeju, Korea (South), August 2016.
[13] S. Xie, "Sensing of mobile device threat status based on big data and SOM neural network," *Boletin Tecnico technical. Bulletin*, vol. 55, no. 7, pp. 332–340, 2017.
[14] W. Yali, "Human resource allocation and performance analysis based on DEA model," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 754–758, 2017.
[15] M. Zheng, X. Ming, and G. Li, "Dynamic optimization for IPS2 resource allocation based on improved fuzzy multiple

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9854787, 1 page
https://doi.org/10.1155/2023/9854787

Retraction

Retracted: Recognition and Error Correction of Piano Playing Features Based on Filtering Technology

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Fan, "Recognition and Error Correction of Piano Playing Features Based on Filtering Technology," *Security and Communication Networks*, vol. 2022, Article ID 6382743, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6382743, 12 pages
https://doi.org/10.1155/2022/6382743

Research Article

Recognition and Error Correction of Piano Playing Features Based on Filtering Technology

Jing Fan

School of Music and Dance, Fuyang Normal University, Fuyang 236000, China

Correspondence should be addressed to Jing Fan; 200007013@fynu.edu.cn

Received 15 January 2022; Revised 25 February 2022; Accepted 28 February 2022; Published 23 March 2022

Academic Editor: Chin-Ling Chen

The extraction of piano features needs to be based on mature technology, but there is no mature technology for feature recognition of music to ensure the correct rate of feature extraction. This article combines filter processing technology to carry out the research on piano playing feature recognition and builds an intelligent error correction system, and it realizes the feature recognition and error correction of piano playing with the support of filter technology. Moreover, this article conducts experimental research on the system constructed. Through experimental research, it can be seen that the piano performance feature recognition system proposed in this article can effectively recognize the piano practitioner's playing effect and record it for error correction. This intelligent method can effectively improve the piano playing effect.

1. Introduction

Music and its performance are always high-level activities that humans engage in. Compared with language, it belongs to a higher level of communication between humans. It is a bridge for expressing emotions between people. When listening to a song, the audience can understand the connotation of the song and can resonate with the emotion of the author. As the process of this kind of emotional transmission, "performance" plays a vital role, and a good player can experience the emotion of music and pass it to the audience through the instrument. However, a failed performance cannot impress the audience or even make the audience feel disgusted. Although the specific music played can be heard, it cannot give the audience a pleasant feeling [1]. The quality of music performance here is a vague concept, and there are many definitions [2], such as fluency of performance, understanding of musical scores, and ability to control musical instruments. In playing teaching, we often hear the teacher say to the students, "Play with your heart." This is a vague concept. In essence, it is something that cannot be depicted on the performance music score. It can only be explained in abstract language, and then, it is up to the students to grasp it [3]. For example, the score on the music score can describe whether a note is "stressed" or not, but it cannot be expressed as the "absolute physical amplitude" of the sound of each note in different playing environments. This ambiguity and uncertainty have become the factors that distinguish the quality of music performance. This is also the reason why ordinary music players and master performers can give listeners different feelings when playing the same piece.

The process of listening to the sound by the audience: the music heard needs to be analyzed by a person with sensitive hearing, and the sound contained in it is represented by a spectrum. The process of musical instruments forming sound waves: the source of sampling is the waveform of the sound, that is, the recording of the amplitude of the sound wave within a certain sampling frequency, such as CB audio. To perform various analyses on this secret audio data, the desired musical characteristics can best be obtained, which involves a large amount of pattern recognition knowledge. Pattern recognition is still a hot research topic, and there are still many areas that need to be improved. Sound pattern recognition has a wide range of applications in speech recognition, and it has also achieved certain research results in music feature extraction.

This article combines filter processing technology to carry out the research on the recognition of piano playing features and builds an intelligent error correction system to improve the effect of piano practice and piano performance.

This article combines the filtering processing technology to carry out the research on the feature recognition of piano playing, constructs an intelligent error correction system, and realizes the feature recognition and error correction of piano playing with the support of filtering technology. The research shows that the recognition of piano performance features can realize the effective recognition of the performance of piano practitioners and record them for error correction.

The organizational structure of this article is as follows: The first part is to describe the needs and background of the identification of piano playing characteristics and to analyze the research motivation of this article. The second part is to analyze the research status of piano exploration characteristics, summarize the literature, and elicit the research content. The third part is the research on the selection method of piano score data smoothing, which is mainly the algorithm improvement part. The fourth part is based on the algorithm based on the recognition of the piano playing features and the construction of the error correction model based on the filtering nursing technology and verify the effect of the model. The conclusion part is a summary of the research method and research content and an outlook.

The main contributions of this article are as follows: (1). The improved algorithm in this study overcomes the large error of the traditional model system and improves the accuracy of the selection of the piano score data smoothing method. (2). The filtering technology can effectively eliminate the interference of external factors in the recognition of piano playing features, which is important for feature extraction.

2. Related Work

For piano playing training, scholars at home and abroad conduct research through three aspects: the brain mechanism of piano playing training, skill level assessment, and auxiliary training system. Kiliç [4] studied the difference in the participation of the motor cortex of the brain during the training of piano music playing between professional and nonprofessional piano players. Studies have shown that the involvement of the brain's primary and secondary motor cortex depends on the stage of motor learning and the subject's experience. In the early training stage, the nonprofessional performers showed that more main and auxiliary motor areas of the brain were activated, and the activation time was longer; in the later training stage, the activity of the main motor cortex of the nonprofessional players decreased rapidly, whereas in the professional players, there was no significant decrease in the activity of the main motor cortex of the brain. Hoffman and Novak [5] proposed a convenient and accurate tool to quantitatively evaluate piano performance techniques. They processed the MIDI files generated during piano performance to obtain the playing interval and key-down speed to evaluate the pianist's teaching and rehabilitation monitoring. Jia et al. [6] designed a tactile-guided magnetic permeable keyboard system MaGkeys, which combined with the audition learning mode to assist learners in piano playing training. Waldron et al. [7] designed a wearable wireless tactile piano teaching system, PianoTouch, which installed five small vibration motors and a Bluetooth module in a glove, and subjects felt the sound from playing notes while listening to the piano music. The vibration of that finger, the results of the study, showed that subjects wearing PianoTouch with built-in tactile perception had better playing effect when playing piano music. Zhang and Tao [8] extracted MIDI information to represent the rhythm, expressiveness, and musicality of performance through MIDI equipment and then integrates neural network to evaluate performance and finally establishes an evaluation system for piano performance and has a good test effect.

Gun [9] conducted relevant experiments on the speed of the piano keys and the spectrum analysis of the keys, and the key touch methods are divided into "raise the finger and press the key quickly" and "paste the key and press the key slowly." The author believes that the final difference lies in the key touch. The strength of the button is different, and it has nothing to do with whether the finger is raised because these two key touch methods only involve strong and weak strengths. In the experiment, the speed of touching the keys is calculated by measuring the time between the fingers touching the keys and the hammers hitting the strings, and the distance between the keys and the hammers, and finally calculating the speed. The measured speed of piano playing is the average speed, and the speed of the finger when touching the key is not a uniform movement but an acceleration movement. The speed obtained in this way cannot reflect the true state of the finger touching the key [10]. In the spectrum analysis comparing different key touch methods, the author analyzed the time-domain and spectrum characteristics of the sound waveform and the part about spectrum characteristics. There are many studies on the relationship between the material, performance, specifications, and the change of each parameter of the action machine and the timbre spectrum, but this has little connection with the research topic of this article—the key touch method and timbre [11]. Lian [12] gave a detailed explanation of the essentials of vertical and horizontal finger touch actions. Kim et al. [13] divided touch keys into finger force touch keys, hand force touch keys, and arm drop according to the force of the touch keys, touch keys, and full arm touch keys. Kaplan and Haenlein [14] believed that the color change of the piano sound is determined by the different strength of the hammer hitting the strings. The greater the strength, the sharper the sound. This is because the upper overtone is generated, and the appearance of the upper overtone is due to the auxiliary vibration of the string. When the power is small, the upper overtone disappears, and the sound becomes soft. Reyes [15] believed that the touch technique of the finger will lead to the change in the overtone above the tone. The greater the distance between the fingers and the keys, the more obvious the effect of the upper overtones, and the overtones above the keys will be weakened or even disappear when the keys directly touch the keys. Jones [16] believed that the fuller and brighter the sound, the more beautiful it sounds. The player can control the tone by controlling the number of overtones at the touch of a key. Under the premise of maintaining the same key height, if you increase the touch area (folding your fingers increases the touch area), the number of upper overtones will be increased, and the timbre will also be changed.

Aithal and Aithal [17] studied the instantaneous motion state of the piano action machine through different key-touching methods and different key-touching strengths. Accelerometers are used to measure the speed of movement of the keys and hammers, and pickups are used to pick up sound signals. The key touch methods they chose were pressed touch and strike touch, and the intensity ranged from weak to strong. The measurement data include the acceleration of the finger key, the acceleration of the hammer string, the time the key is in contact with the key bed, the maximum speed of the hammer, and so on. The conclusion is that the length of time between the finger touching the key and the hammer hitting the string can be roughly regarded as the maximum speed of the hammer after calculation, and this speed is very different between the two key touching methods [18]. But when the two keystrokes touch the keys with little force, the difference in speed becomes smaller.

3. Selection of Piano Score Data Smoothing Method

3.1. Linear Smoothing of Filter Data. The polynomial least-squares fitting method uses a polynomial of degree n and $2m+1$ data points to perform smooth fitting section by section to achieve the purpose of smoothness. This method takes each data of the γ-wave spectrum original data as the center and takes m data (total $2m+1$) to the left and right sides for smoothing. The corresponding data channel value coordinate is $(-m. -m+1. -m+2 \ldots -1.0.1 \ldots m-2.m-1.m)$, and the corresponding channel value count is $y-m, y-m-1, y-m-2 \ldots y-1.$ $y_0.y_1 \ldots y_{m-2}.$ $y_{m-1} \cdot y_m°$. This article uses an n-degree polynomial to fit these data (n is less than $2m+1$):

$$\widetilde{f}_i = \sum_{k=0}^{n} b_{nk} i^k = b_{n0} + b_{n1} i + b_{n2} i^2 + \cdots + b_{nn} i^{kn}. \qquad (1)$$

According to the requirements of the nonlinear least square method, the difference between the actual observation value y_i and the polynomial calculation value f_i should be the smallest, that is,

$$\frac{\partial}{\partial n_{\mathrm{nk}}}\left[\left(\sum_{i=-m}^{m}\sum_{k=0}^{n} b_{nk} i^k - y_i^2\right)^2\right] = 0. \qquad (2)$$

For bn_0,

$$\frac{\partial}{\partial n_{\mathrm{n0}}}\left[\left(\sum_{i=-m}^{m} b_{n0} + b_{n1} i + \cdots b_{nn} i_n - y_i^2\right)^2\right] = 0. \qquad (3)$$

For general b_{ni}:

$$\sum_{k}^{n} b_{nk} \sum_{i=-m}^{m} i^{k+r} = \sum_{i=-m}^{m} y_i i^r. \qquad (4)$$

For $2m+1$ data points, a polynomial of degree n is used as the least squares fitting formula, and the general formula is derived:

$$\widetilde{y}_l = \frac{1}{K_b}\sum_{j=-m}^{m} A_j A y_{i+j}. \qquad (5)$$

According to formula (5), the smoothing coefficient j, A and the normalization constant b, K are calculated according to the corresponding filter. We use the Savitzky-Golay filter to obtain the coefficient calculation formula of the second or third-degree polynomial spectrum smoothing formula [19]:

$$f_{w,j} = \frac{1}{w}\left[1 + \frac{15}{w^2-4}\left(\frac{w^2-1}{12} - j^2\right)\right] w = 5, 7 \ldots 2m+1. \qquad (6)$$

According to formula (6), the five-point smoothing formula of cubic polynomial used in this article is given as follows [20]:

$$y_i = \frac{1}{35}\left(-3y_{i-2} + 12y_{i-1} + 17y_i + 12y_{i+1} - 3y_{i+2}\right). \qquad (7)$$

According to formula (7), the table of smoothing coefficients of the third-order smoothing method at 5, 7, 9, and 11 points can be obtained as shown in Table 1.

The five-point smooth first derivative formula of the cubic polynomial used in this article is

$$\mathrm{data}_i' = \frac{1}{12}\left(\mathrm{data}_{i-2} - 8\times\mathrm{data}_{i-1} + 8\times\mathrm{data}_{i+1} - \mathrm{data}_{i+2}\right). \qquad (8)$$

From the actual effect point of view, the quality of the spectral line smoothing effect also has a great influence on the peak finding results. When the spectral line is not smoothed, the peak position cannot be found at all. If the smoothing effect is not good, misidentification and missed peak identification will occur. The derivative method is effective in identifying single peaks, strong peaks, and weak peaks, but there are problems in the identification of overlapping peaks. From the sensitivity point of view, the sensitivity of the first-order peak search is the highest, and the second-order is the second. When searching for peak positions, some additional conditions are added to enhance the accuracy of peak searching. The flowchart is shown in Figure 1.

The additional conditions are as follows:

(1) The distance N between the two negative poles of the first derivative is between 1 time the half-height width and 4 times the half-height width

(2) The count at the peak position track value should be greater than 4

(3) It cannot be a mutation (a mutation is not allowed within a few tracks on the left and right sides of the peak)

(4) The slopes on the left and right sides of the peak position are set with corresponding thresholds

The peak positions found by the derivative peak finding method are all integer channels. For the needs of qualitative

Table 1: Parameter table of least squares fitting method.

Smoothing points	5	7	9	11
k_b	35.0	21.0	231.0	429.0
y_1	17.0	7.0	59.0	89.0
$y_{i\pm1}$	12.0	6.0	54.0	84.0
$y_{i\pm2}$	−3.0	3.0	39.0	69.0
$y_{i\pm3}$		−2.0	14.0	44.0
$y_{i\pm4}$			−21.0	9.0
$y_{i\pm5}$				−36.0

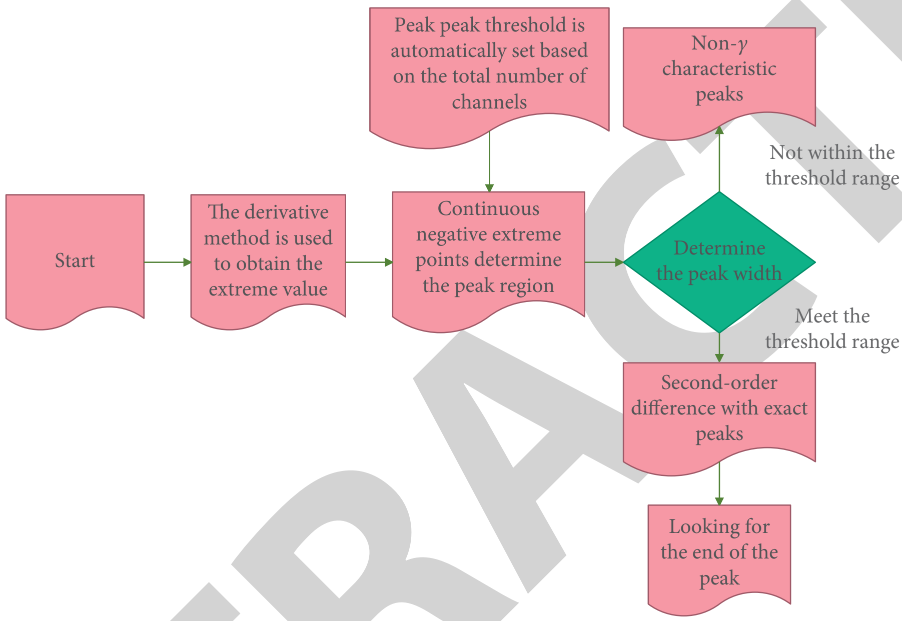


Figure 1: Flow chart of derivative peak finding.

analysis, the peak positions should be obtained more accurately. This article uses the second-order difference polynomial to accurately calculate the peak positions. First, the algorithm performs fitting at the three points $i-1$, i, and $i+1$ around the peak position i found, $\text{data} = a + b \times i + c \times i^2$, and the extreme point is the peak position [21]:

$$ps = -\frac{b}{2c} = i + \frac{1}{2}\frac{\text{data}_{i+1} - \text{data}_{i-1}}{2\text{data}_i - \text{data}_{i+1} - \text{data}_{i-1}}. \qquad (9)$$

In actual measurement, the peak width will be different for different detectors, different total channel values, and energy intervals. Therefore, in order to make the peak search have good adaptability, for different detectors and total track values, set the corresponding peak width interval. After inputting the parameters into the software, even if the detector is changed, it can be directly measured without manual modification on-site, and the algorithm's adaptability is enhanced. The specific parameters are shown in Table 2.

There will be deviations between the actual calculated peak position and the theoretical peak position, so an energy window needs to be set. When the energy difference between the peak position energy obtained by the peak finding and the feature peak of the piano audio feature in the piano audio feature library is less than a specific energy threshold, it is considered that the peak may belong to the retrieved corresponding piano audio feature.

3.2. Piano Audio Characteristic Activity Measurement. In the process of measuring piano audio feature activity, in addition to the energy emitted by the piano audio feature itself, there will also be other interference factors that affect the calculation of the piano audio feature content. In order to eliminate interference factors and improve the accuracy of calculating piano audio feature activity, a portable spectrometer is required to automatically identify and remove filter processing counts other than the full-peak area.

The current background subtraction methods are mainly divided into two categories: one is the background subtraction method of the characteristic peak area: this type of method selects a section of the all-powerful peak area to be analyzed and selects the specific peak area according to the actual characteristic parameters, such as peak shape and peak width. The background subtraction method using the background subtraction method is used to calculate the background to obtain the background count and net count of each track. The other type is the background subtraction method for the whole spectrum: this method obtains the

Table 2: Corresponding peak widths of different total channels.

Total number of channels (LaBr3 (Ce))	1024	2048	4096	Total number of channels (NaI (T1))	512	1024
Energy window width	16.0–72.0	35.0–100.0	49.0–170.0	Energy window width	15.0–60.0	28.0–114.0

entire spectrum by calculating the background counts of all the trace values of the entire spectrum and then subtracting the corresponding background counts from the entire spectrum one by one (net count of lines). At present, most spectral analysis software adopts the background subtraction method of characteristic peak area. Combined with existing research and experimental analysis, the results shown in Figures 2–4 are obtained.

3.2.1. Linear Filter Processing. This method uses the coordinates of the left and right boundary track addresses of each full-wave peak or overlapping peak as parameters to fit a straight line. The counts below the straight line are treated as filtering, and the counts above are treated as net counts. The method is simple to operate. However, the quality of the boundary parameters of the peak area has a great influence on the deduction effect. In particular, the influence of Compton scattering or other rays will result in poor filtering processing deduction. The straight line is used as the baseline to deduct the filtering process, which is susceptible to other scattering effects and has poor anti-interference ability. Generally, the linear filtering process generally only has a better effect when it is used at the full peak of the single energy, as shown in Figure 2.

3.2.2. Step Filter Processing. If the filtering processing of the low-energy end of the full-wave peak is much higher than the filtering processing of the high-energy end, a step function can be used to represent the filtering processing count of the overlapping peaks, as shown in Figure 3. The height of the steps is determined by the maximum value and minimum value of the boundary track value. The data under the step function is regarded as the filter processing count of the full-wave peak, and the above count is the net count.

3.2.3. Parabolic Filtering Processing. In some cases, most of the filter processing counts are on the measured spectrum, as shown in Figure 4. At this time, there will be a fast-rising filter processing in the low-energy part of the full-wave peak. The parabolic filtering process uses the least square method to fit the data, and the count calculated by the submethod is lower than the high-energy end and lower than the straight-line filtering process.

3.2.4. New Filter Processing Deduction Method. This method fits the filter processing with a function, which is regarded as filter processing. Currently, the commonly used methods include SNIP filter processing method, Fourier filter processing method, wavelet transform method, and so on. The SNIP method is a widely recognized and applied method abroad.

This article improves the algorithm and sets the adaptive parameters according to the voice recognition requirements used. In the filtering process and subtraction process, the natural logarithm method can get good results when the count rate is high, and the square root method can also get good results when the count rate is low. Therefore, the LIS logarithmic transformation can be performed on the spectrum, and then the filter processing can be removed by the SNIP method, and finally, the LIS inverse transformation can be performed to obtain the filter processing count. The specific steps of the SNIP algorithm are as follows:

(1) The algorithm first uses the LLS operator to transform each count of the spectrum:

$$v_i = \ln[\ln(\sqrt{y(i)+1} + 1) + 1]. \quad (10)$$

Among them, i is the corresponding channel value, $y(i)$ is the count corresponding to the channel value, and v is the result retention vector. This can compress the range of the count rate and, at the same time, enhance the sensitivity of weak peak recognition.

(2) The algorithm performs multiple iterations on the data and replaces the original value with the obtained value:

$$v_p(i) = \min\left\{\mathrm{v}_{\mathrm{p}-1}(\mathrm{i}), \frac{1}{2}\left[v_{p-1}(i+p) + v_{p-1}(i-p)\right]\right\}, \quad (11)$$

m is the total number of iterations, and p is the p-th iteration. In the p-th iteration, the algorithm takes the smallest value of $v_{p-1}(i)$ and $1/2[v_{p-1}(i+p) +$ $\$\$v_{p-1}(i-p)$ as the value of $v_p(i)$. Ryan performs 24 iterations on the algorithm and shrinks the window size by 1/2 in the last 8 times to eliminate possible oscillations in the results.

(3) The algorithm gets the filter processing count:

$$\mathrm{b}(\mathrm{i}) = \exp\left(\exp\ (v(i)-1)^2\right), \quad (12)$$

$b(i)$ represents the count of the filtering processing of the full peak.

(4) The net peak area of the full-wave peak is given as

$$n(i) = y(i) - b(i). \quad (13)$$

The width of the energy window is W, the number of iterations is m, and the relationship between w and m is $w = 2m + 1$. It can be seen that the selection of the full-wave crest boundary has a great influence on the deduction effect of the filtering process. Figure 5 shows the deduction effect of the filtering process with different iteration times. It can be

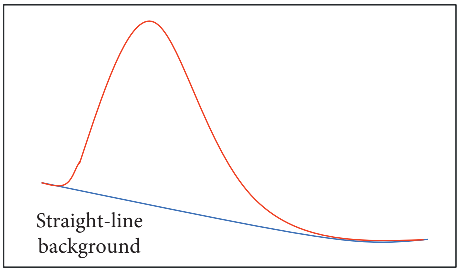


Figure 2: Linear filter processing.

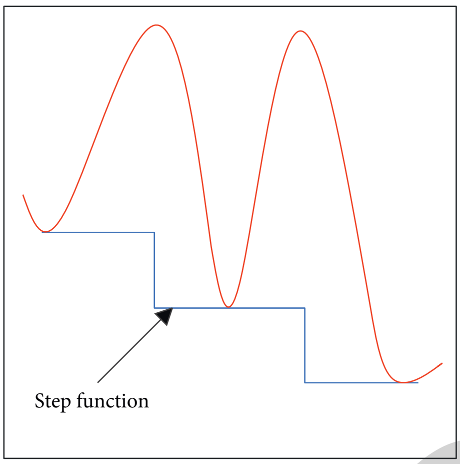


Figure 3: Step filter processing.

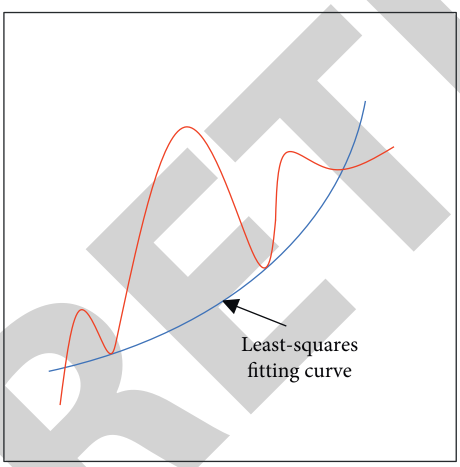


Figure 4: Parabolic filtering processing.

seen from Figure 6 that when the number of iterations is about 0.5 w, the deduction effect is similar to linear filtering.

3.3. SNIP Algorithm Improvement. This article improves the algorithm based on the original SNIP algorithm. First, the method of decreasing the width of the energy window from m to 1 is used to iteratively calculate formula (11). Second, in selecting the number of iterations, the value of m is determined by the difference between the heights of the full-wave crests. Finally, the fourth-order filter function is used

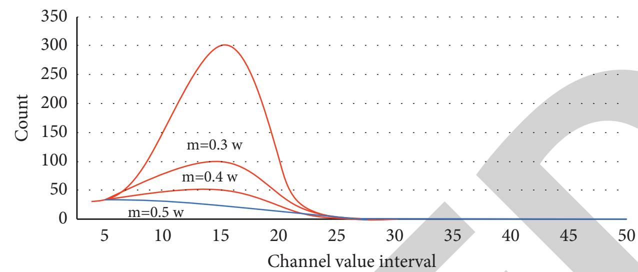


Figure 5: SNIP filter processing deduction effect.

for iteration instead of the second-order filter function, as shown in formula (14):

$$
\begin{aligned}
a &= y(i);\\
b &= \frac{[y(i-m)+y(i+m)]}{2}\\
c &= \frac{[-y(i-m)+4y(i-m/2)+4y(i+m/2)-y(i+m)]}{6}\\
y'(i) &= \min[a,\ \max(b,c)].
\end{aligned}
\tag{14}
$$

After comparative analysis of the measured data, it is found that selecting 1/3 of the height difference between the left and right borders of the full-wave crest is better for iteration, as shown in Figure 6(a). Figures 6(b) and 6(c) show the deduction effects of linear filter processing and step filter processing.

For heavy peaks (independent feature peaks), SNIP filtering processing method cannot be used to directly subtract. First, this article uses the found boundary to perform linear filter processing and subtraction and then performs Gaussian fitting on the two peaks after subtracting the filter processing count. After that, this article redetermines the boundary of the two peaks based on the fitted function and the peak position and half-height scale information of the two peaks. Finally, this article uses the SNIP filter processing subtraction method to perform filter processing subtraction to obtain the filter processing count, as shown in Figures 7(a) and 7(b).

In spectrum analysis, peak area determination is the most important step of quantitative identification. This method reflects the content and activity of piano audio features by calculating the peak area count of feature peaks. Obviously, the accuracy of the peak area calculation directly affects the performance of the piano audio feature activity calculation usually. There are two types of methods for calculating peak areas. The first type is counting addition. This method only needs to select the peak area that needs to be calculated and add the counts in the peak area to get the value of the peak area. It is usually used when there is no single peak that interferes with each other. The calculation results also have corresponding accuracy. The second is called function fitting method. This method uses a function to fit the full peaks (such as Gaussian function fitting, polynomial fitting, and the like) and then integrates the fitted

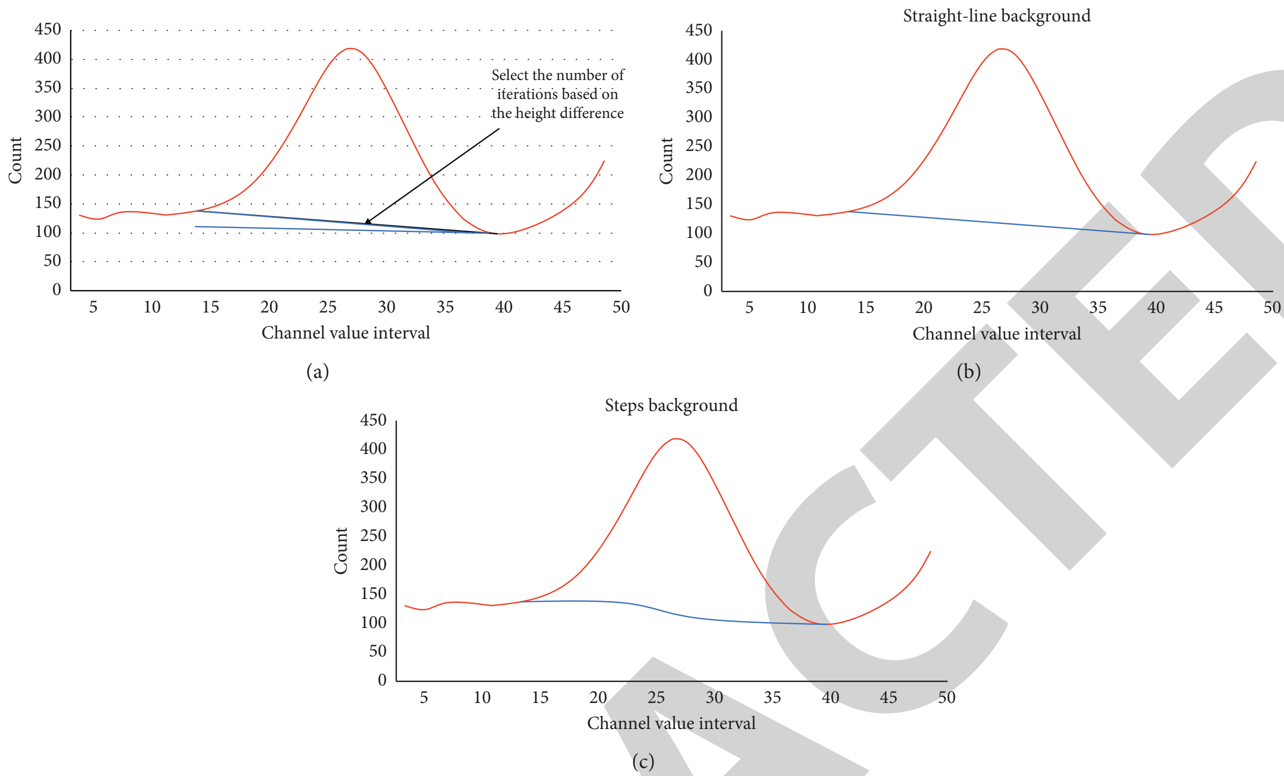


FIGURE 6: Test data statistics. (a) Effect diagram of SNIP filtering processing subtraction method. (b) Linear filter processing. (c) Step filter processing.

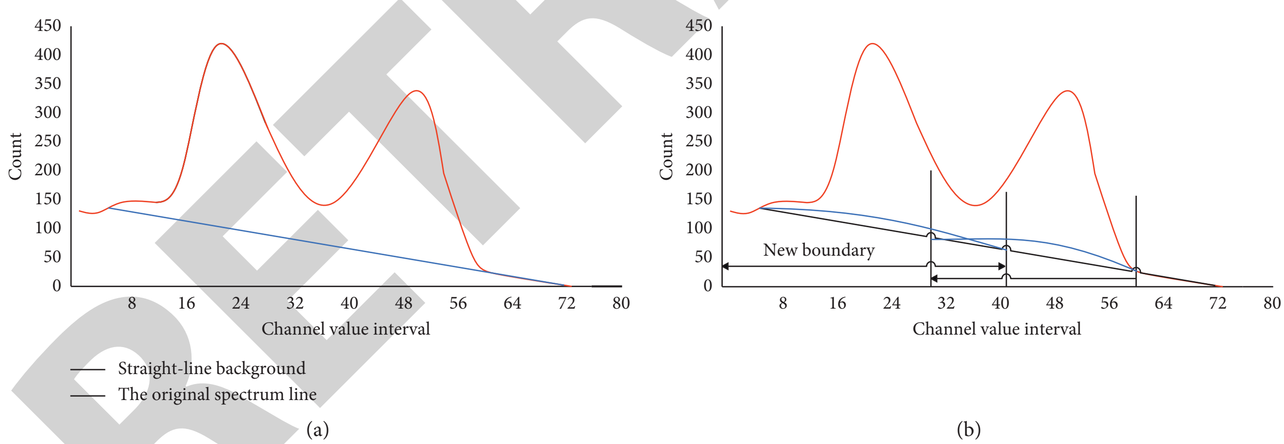


FIGURE 7: Filter processing count. (a) Linear filter processing. (b) Redetermine the boundary for SNIP filtering processing deduction.

function to obtain the corresponding peak area and half-height width and other corresponding parameters. This method needs to be run on a computer and is usually used in the fitting analysis of heavy peaks, and the calculation results also have relatively good accuracy. For portable smart devices, too complex fitting functions cannot be used.

The counting addition method can be divided into full-peak area method, Covell method, Wasson method, Sterlinski method, and Quittner method according to the difference of leakage filtering processing and boundary selection method.

3.3.1. Full-Peak Area Method. This method is also called the TPA method, which uses the peak position and boundary found, adds up all counts within the left and right boundaries of the peak, and deducts the filtering process to obtain the net count. Filter processing can choose the corresponding method according to the actual situation. In the process of determining the peak area, there are various reasons that can cause errors. The error of the full-peak area method mainly comes from two aspects. First, which method is used for the filtering method to reduce the error is to be determined according to the actual spectrum. Therefore, when the computer automatically performs the filtering process, it is difficult to adaptively select the corresponding method, and there is no filtering process deduction method that can well adapt to all filtering processes. At the same time, the filtering process is not limited to the environmental filtering process, but it also includes the improvement of the filtering process caused by other high-energy interference piano audio features of the Compton platform. Therefore, this method is susceptible to the inaccuracy of the filtering processing deduction. For a single peak, this article uses this method to calculate the area of the full peak, and the method of subtracting the filtering processing uses the improved SNIP method for filtering processing subtraction, which has achieved good results. The test proves that this method is obviously better than the linear filtering processing subtraction method. The second is to count statistical errors, and the formula for statistical errors is

$$\sigma_N^2 = N + \frac{1}{2}(L - R - 1)B. \tag{15}$$

According to formula (15), it can be seen that the variance of the peak area is related to the full-peak area N and the filter processing count B. However, the coefficient factor of B is 0.5 (L–R–1) and the coefficient of N is 1. It can be seen that the influence of the filtering process on the error is far greater than the influence of the peak area. Therefore, it is very important to choose a good filtering method. Therefore, other methods have been developed. Although TPA has higher requirements for filter processing and deduction, it uses all pulse counts within the peak and is minimally affected by peak drift and resolution. Therefore, it is still widely used in the calculation of single peaks.

3.4. Full-Peak Fitting. The function fitting method uses a known function to describe the peak based on the measured peak area data and calculates all the relevant parameters in the function (such as half-height and peak height). Then, the peak area can be calculated by integration. Commonly used function fitting functions include Gaussian function fitting, least square fitting, polynomial fitting, and so on.

In this article, Gaussian fitting is used to fit the full peaks of heavy peaks. The basic process is as follows:

A peak can be described by a Gaussian function, that is, the relationship between each track count $y(x)$ in the peak area and the track number X is given as

$$y(x) = y_0 e^{-(x-x_0)/2\sigma^2}. \tag{16}$$

In the formula, x_0 is the number of peak center channels, and y_0 represents the peak center channel count, namely, $y_0 = y(x_0)$. σ is a parameter describing the width of the peak distribution (root mean square error), and its relationship with the half-maximum width (FWHM) is

$$\text{FWHM} = 2\sqrt{2\ \ln 2}\sigma = 2.36\sigma. \tag{17}$$

According to the peak information, this article first analyzes the peak to find whether it is a single peak or a heavy peak (the peak position is separated, and the peak area partially overlaps). For a single peak, the algorithm directly calculates the total peak area by counting and adding and then uses the SNIP method to subtract the filtering process to obtain the net peak area of the full peak. For heavy peaks, it is divided into the following three steps:.

(1) The overlapping peaks are first subjected to linear filtering processing to subtract, and the count of the peak area is obtained. The linear filter processing deduction is shown in Figure 8(a).
(2) The algorithm performs Gaussian fitting on the two peaks deducted from the linear filtering process. The peak fitting method on the left is Gaussian fitting using the noncoincident part on the left side of the peak position. Similarly, the peak on the right is also Gaussian fitting using the noncoincident part on the right of the peak. After that, the algorithm uses the FWHM scale function to find the FWHM of the peak and uses the FWHM and the peak position track value to redetermine the left and right boundaries of the peak, as shown in Figure 8(b).
(3) The algorithm uses the redetermined boundary to perform SNIP filtering processing deduction, and then, it uses the fitted total count to subtract the filtering processing count to obtain the net count of the full peak.

4. Recognition and Error Correction of Piano Playing Features Based on Filtering Technology

The system in this article is implemented on the Matlab platform. This article combines the previous algorithm to identify the features of piano playing and builds an intelligent piano-assisted practice error correction system. As shown in Figure 9, the overall architecture of the system consists of three parts. The first one is the web backend, including business logic code, object storage, database, and the like. The second part is the web front end, including the page displayed on the WeChat official account and the page displayed on the PC. The third part is the iPad client, and performance-related services are implemented in this part.

With the support of filtering technology, the system can realize the feature recognition and error correction of piano playing. This article conducts experimental research on the

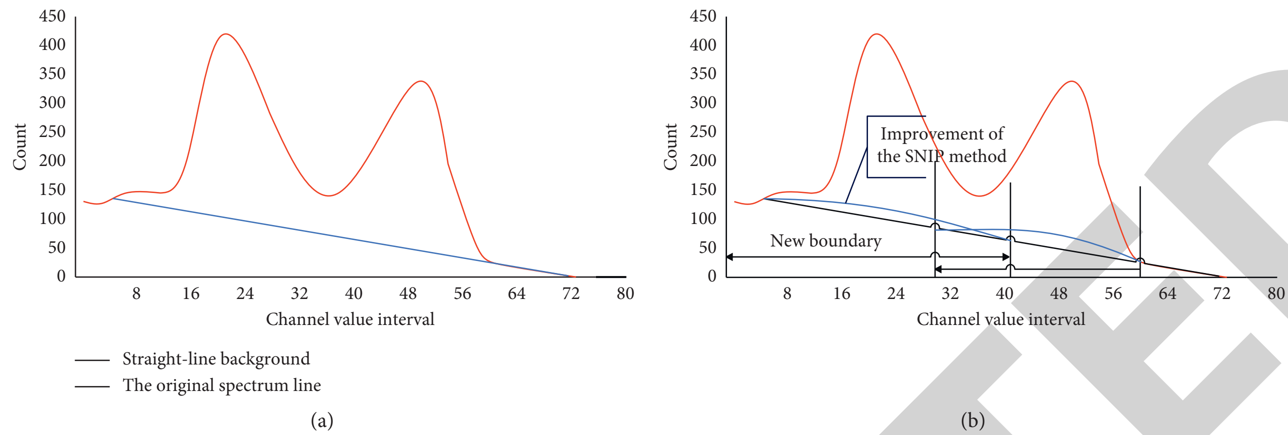


FIGURE 8: Value determination. (a) Straight-line filter processing deduction. (b) Heavy peak fitting and filtering processing.

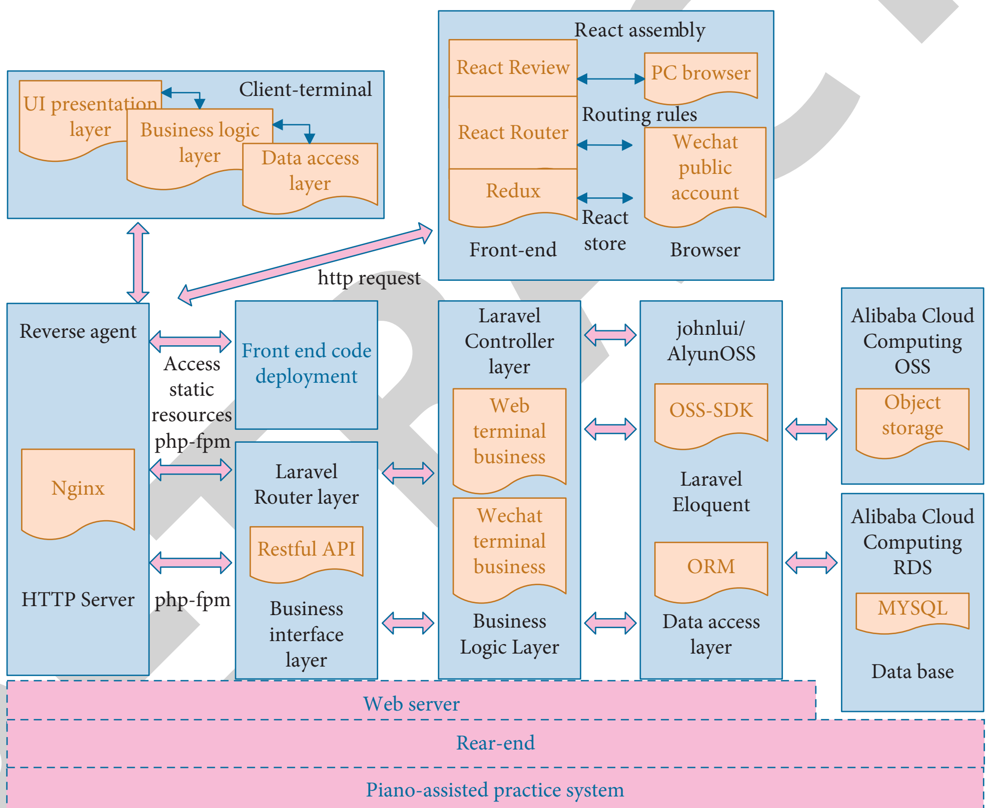


FIGURE 9: System overall architecture.

system constructed, and studies the difference between keystrokes and keystrokes. From experience, the difference between keystrokes and keystrokes basically does not exist. Therefore, there is no difference between the two settings, and on this basis, the waveform processing is carried out through the system of this study. The frequency spectrum of “hit the key “and “press the key” is shown in Figure 10.

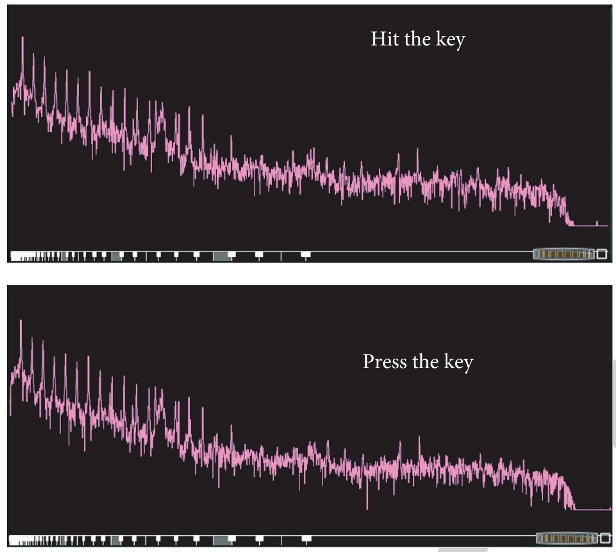


Figure 10: The frequency spectrum of "hit the key" and "press the key".

Table 3: Piano playing feature recognition and error correction effect evaluation.

Number	The method of this paper	The method of literature [3]	Number	The method of this paper	The method of literature [3]
1	82.90	80.07	27	86.08	77.90
2	80.75	72.38	28	81.09	76.24
3	88.27	82.82	29	84.37	80.01
4	80.58	78.61	30	90.48	89.26
5	81.71	70.18	31	82.00	71.18
6	88.24	78.75	32	82.89	71.95
7	80.16	77.73	33	90.09	82.13
8	90.75	88.70	34	92.10	87.69
9	91.22	85.61	35	85.95	83.77
10	85.80	78.82	36	88.53	87.44
11	91.89	78.71	37	87.98	86.15
12	79.50	72.80	38	84.50	77.42
13	89.40	88.74	39	92.61	92.55
14	85.62	80.85	40	86.49	80.32
15	82.58	75.54	41	92.47	82.31
16	90.61	87.23	42	90.44	88.24
17	90.47	88.16	43	81.27	76.76
18	86.35	76.21	44	84.73	83.87
19	86.59	82.64	45	86.36	78.63
20	88.00	82.19	46	83.11	78.75
21	82.85	77.34	47	86.64	77.04
22	84.99	78.66	48	90.56	80.01
23	80.90	73.17	49	80.73	71.85
24	79.18	76.05	50	81.54	71.72
25	92.81	82.81	51	84.84	79.43
26	85.14	82.66	52	89.37	78.95

On the basis of the above research, the piano playing feature recognition and error correction effect evaluation of the system constructed are carried out through experimental analysis of multiple sets of playing. The research content of this article is compared with the literature [3], and the results are shown in Tables 1 and 2, and the results are shown in Tables 3 and 4.

From the above research, it can be seen that the piano playing feature recognition and error correction system based on filter processing technology proposed in this study

Table 4: Error correction effect evaluation.

Number	The method of this paper	The method of literature [3]	Number	The method of this paper	The method of literature [3]
1	81.97	70.31	27	87.31	79.53
2	79.03	70.69	28	78.80	73.18
3	79.06	70.55	29	88.18	81.78
4	75.83	65.67	30	74.86	67.64
5	74.80	71.55	31	74.97	70.68
6	78.92	70.37	32	86.11	76.11
7	89.92	87.51	33	74.62	66.45
8	72.84	72.31	34	84.20	81.69
9	88.45	76.32	35	77.63	76.31
10	77.09	69.74	36	72.55	66.80
11	83.05	77.29	37	89.22	78.93
12	76.95	75.32	38	88.01	76.93
13	81.26	73.56	39	85.24	83.59
14	82.82	73.32	40	78.14	71.21
15	73.06	67.63	41	75.35	75.16
16	87.15	81.53	42	81.78	80.68
17	89.53	77.69	43	84.25	72.41
18	72.35	69.21	44	75.77	68.12
19	80.06	78.78	45	73.31	70.88
20	85.01	78.65	46	85.64	81.70
21	78.70	67.05	47	79.16	76.70
22	83.16	81.80	48	73.26	62.75
23	75.73	75.13	49	80.78	79.63
24	87.45	81.26	50	77.39	69.94
25	72.06	68.33	51	76.50	66.66
26	79.93	75.30	52	84.95	81.26

has good results, so the system can be used as an aid in subsequent piano playing.

5. Conclusion

An isolated note and a chord cannot make up music, but when multiple notes are combined, they can show a beautiful piece of music in a certain rhythm. The good or bad grasp of factors such as tone, chord, melody, and so on when playing has become the basis for judging whether a piece of music is good to hear. The features of piano performance directly reflect the emotional features of music. Today, with the rapid development of information technology, the type and amount of information that people obtain is advancing by leaps and bounds. As the main carrier of current information, multimedia technology has attracted more and more attention. Audio is one of the important forms of multimedia information. Through the recognition of piano performance characteristics, the effective recognition of piano practitioners' playing effects can be realized, and errors can be recorded and corrected. Moreover, this intelligent method can effectively enhance the effect of piano playing. Therefore, this article combines the filter processing technology to carry out the research on the recognition of piano playing characteristics and builds an intelligent error correction system to improve the effect of piano practice and piano performance.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This study was sponsored by 2021 Social Science Foundation of Anhui Province: modern expression of wind and percussion music and its humanistic narration in Anhui province (AHSKY2021D118).

References

[1] G. Bedi, G. K. Venayagamoorthy, R. Singh, R. R. Brooks, and K.-C. Wang, "Review of internet of things (IoT) in electric power and energy systems," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 847–870, 2018.

[2] I. Bisio, A. Delfino, A. Grattarola, F. Lavagetto, and A. Sciarrone, "Ultrasounds-based context sensing method and applications over the internet of things," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3876–3890, 2018.

[3] A. Chamberlain, M. Bødker, A. Hazzard et al., "Audio technology and mobile human computer interaction," *International Journal of Mobile Human Computer Interaction*, vol. 9, no. 4, pp. 25–40, 2017.

[4] D. B. Ç. Kiliç, "Pre-Service music teachers' metaphorical perceptions of the concept of a music teaching program[J]," *Journal of Education and Learning*, vol. 6, no. 3, pp. 273–286, 2017.

[5] D. L. Hoffman and T. P. Novak, "Consumer and object experience in the internet of things: an assemblage theory approach," *Journal of Consumer Research*, vol. 44, no. 6, pp. 1178–1204, 2018.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9765420, 1 page
https://doi.org/10.1155/2023/9765420

Retraction

Retracted: Application of Digital Image Processing Technology in the Remote Interactive Art Teaching System

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] G. Meng, "Application of Digital Image Processing Technology in the Remote Interactive Art Teaching System," *Security and Communication Networks*, vol. 2022, Article ID 6419117, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6419117, 12 pages
https://doi.org/10.1155/2022/6419117

Research Article

Application of Digital Image Processing Technology in the Remote Interactive Art Teaching System

Gang Meng

Academy of Fine Arts, Hexi University, Zhangye, Gansu 734000, China

Correspondence should be addressed to Gang Meng; 2001618189@hxu.edu.cn

Received 23 January 2022; Revised 21 February 2022; Accepted 28 February 2022; Published 17 March 2022

Academic Editor: Chin-Ling Chen

Modern distance teaching activities have occupied an important position in current education and teaching activities. This research mainly discusses the application of digital image processing technology in remote interactive art teaching systems. In order to avoid complicated calculation and a huge amount of calculation, reduce the difficulty of calculation, and improve the correction speed of distorted image and for qualitative recognition and correction of a large background and small target (or medium and small target) digital images (video is regarded as a special form of continuous image, method-like) that appear in a large number of remote teaching activities, this paper proposes a direct look-up method based on coordinate interpolation, that is, to establish the approximate correspondence between the image coordinate positions before and after the distortion and build a reference model for comparing the two image coordinates through experimental data. The system design includes overall design, functional module design, and database design. To realize the function of distance education, it must have characteristics different from ordinary networks and the most important thing is that its transmission bandwidth should be large. Because the distance education platform needs to be able to transmit the teaching text, hear the teacher's voice, see the teacher's teaching image content, and even see the carefully designed animation, the distance education system must have a very wide bandwidth. The database tables are designed using entity attribute diagrams and E-R diagrams. E-R diagram is also called entity-connection diagram, which provides a way to express entity types, attributes, and connections. It is used to describe the conceptual model of the real world. Remote online test management includes test content perform operations such as addition. The art teaching management system designed in the article will make teachers and students no longer be restricted by time and place and can inquire about students' information, choose courses freely, inquire about grades, and understand the teaching plan anytime and anywhere. This research helps to promote the development of remote interactive technology.

1. Introduction

Many traditional teaching methods include lecture, demonstration, listening, memorizing, practice, observation, etc. The prerequisite is to ensure that students actively perceive and understand the knowledge of the subject's purpose, do not perform abstract rote memorization, but can adapt to the situation when solving problems. So as to consolidate and improve the knowledge reserve to be able to cope with various examination questions and so on. As the country pays more and more attention to school education, campus network facilities are becoming more and more perfect. In order to learn and comprehend knowledge better, more conveniently, and more quickly, we should combine advantages and transform the previous traditional teaching methods to the current intelligent teaching methods so that learning efficiency can be continuously improved. Most of the art education focuses on the art itself and focuses on the teaching of skills. All of these severely restrict the development of art education and teaching in high schools. Therefore, the reform of the new art curriculum that meets the requirements of quality education has improved the status and provided space for art education.

In such a big environment and when the school wants to continuously improve course teaching, the construction course website fully demonstrates its necessity. It is a form of education that transmits courses to one or more students outside the campus. It is an education form in which students and teachers, students and educational organizations, teachers and educational organizations mainly adopt a variety of media methods for systematic teaching, educational management, and communication. To establish such a course website, the capital investment is not large, required not too much, and the operation and maintenance of the website are relatively convenient, so the construction of the website is feasible. The intelligent interactive education system has made some achievements in recent years, but the current interactive education system cannot meet this requirement. Therefore, in general, although the interactive education system has made great progress, there are still big problems which affect the modernization process of education and teaching in our country to a certain extent. The development of modern education has important practical significance.

Remote interactive teaching can optimize the classroom structure, improve teaching efficiency, and stimulate students' creative thinking. It has many advantages and functions that other media (such as slides, projection, etc.) do not have or are not fully equipped. It has become a development trend of modern teaching and has a profound influence and great significance on educational concepts, teaching methods, and teaching organization forms. This paper discusses the construction process of an auxiliary. Teaching provides students with learning anytime, anywhere, so that life-long learning becomes possible. Students learn in a brand-new way, which can achieve certain educational goals well in teaching. Therefore, modern teachers should give full play to the educational functions and resource advantages of the network and strengthen interactive teaching activities.

The main research structure proposed in this paper is as follows:

Section 1 briefly introduces the research background, research status at home and abroad, research significance, and research content.

Section 2 evaluates previous related work.

Section 3 is an overview of related technologies. This section provides a detailed overview of the techniques and techniques to be used in this study.

Section 4 is the demand analysis and design of the art teaching and research system. According to the characteristics of art education and teaching, this section analyzes the current requirements of art education and teaching for the system and realizes the analysis of the art education and teaching system through functional demand analysis and module analysis.

Section 5 is the realization of the art teaching and research system. Through the analysis of the presentation layer, control layer, logic layer, and persistence layer, the process and steps of the art teaching and research system are described and designed in detail. And it analyzes the important strategies and algorithms adopted in the current online art teaching system and finally explains how the paging technology is realized.

Section 6 summarizes the research content and results of this paper and proposes future research directions by analyzing the shortcomings of this research.

2. Related Work

Although traditional online teaching can also allow students to watch the teacher's lecture, this information receiving mode is nonreal-time and one-way for students. Students will feel bored and passive, resulting in low classroom efficiency. To fully achieve the goal of curriculum website construction, it needs long-term hard work. In teaching, we must continuously integrate teaching practice, focus on improving the quality of teaching, and constantly adopt new measures website and improvement of teaching quality. Kołodziejczak and Roszak believe he introduced the knowledge level [1]. Shields believes that more and more students around the world are participating in distance education courses. Representative participation and pedagogical elements meet the requirements of institutions and industries. For this reason, it is essential to design contemporary courses to ensure these results are achieved. He uses narrative and comprehensive methods to promote understanding. He regards the role of educators as conductors, technicians, and choreographers. Finally, he proposed a triad consisting of pedagogy, technology, and participatory learner community as the basis for ensuring that the curriculum conforms to contemporary practice [2]. Harsasi and Sutawijaya believe that the online learning system has become the main requirement for the implementation of the learning process, which is mainly reflected in Indonesia [3]. Vasilevska et al. believe that the distance learning environment with different methods has become one of the most studied paradigms [4]. Abdullah NA believes that in addition to the traditional model, Pakistan's teacher education programs are also provided through online and distance education. He chose the students who participated in the teaching practice module in the fall semester of 2018 as a sample. Data sources include teaching plans, lectures, administration, and extracurricular work made, teachers' teaching has improved. The teacher's personal visit is very important for verifying and evaluating actual classroom teaching [5]. For a distance teaching system, the difference of the examination subsystem determines the success of the teaching system to a large extent, while for the examination subsystem, the degree of differentiation of the test paper directly determines the pros and cons of the examination subsystem.

3. Interactive Remote Teaching System Design

3.1. Remote Interactive Art Teaching System. After a class is finished, a teaching video is formed and uploaded to the traditional network platform of the distance education organization so that those students who can not participate in the classroom in real time due to various special reasons can watch it. The form of interaction can be text, voice, or even video. Because this kind of teaching video does not only carry out one-way knowledge infusion like traditional teaching media but two-way information interaction, the learning effect of students can be much better. The network teaching system oriented to the fine arts profession studied in this paper is mainly aimed at the teacher-side teaching function that can realize the virtualized classroom environment. Secondly, it is planned to restore the classroom environment for the coursework system, and the system saves and prints the lecture information to provide teachers with an independent examination platform. The system structure diagram is shown in Figure 1. The website design should not only focus on the needs of the initial users of the system but should conduct a more extensive analysis to determine more comprehensive needs. Servlet (Server-Applet) is the abbreviation of JavaServlet, which is called a small service program or service connector. It is a server-side program written in Java and has the characteristics of being independent of platform and protocol. Its main function is to browse and generate data interactively and generate dynamic Web content. From the perspective of system architecture, Servlet greatly reduces the complexity of application construction. From the perspective of website planning, the website will not become simple because of Servlet. But what Servlet brings is better reuse and maintenance, which makes development more flexible so that more work can be put on-demand analysis and implementation and system debugging. The system is mainly divided into six subjects and three connections. The six subjects are students, teachers, administrators, homework, latest information, and friendship links. The relationship between students, teachers, and administrators is management, and the relationship between students, teachers, administrators, and homework is also management. The relationship between students and the latest information and friendship links is browsing, and the relationship between teachers and administrators and the latest information and friendship links is management [6, 7].

The input and output devices are mainly for drawing boards and input pens, and their main functions are similar to the basic tools of using Windows drawing boards. But it is mainly an external input device, and the input process mainly uses different pressure sensing to obtain the length and width of the unit, with a corresponding pressure pen as an aid. It can simulate the feeling of painting in daily life in real time. Mainly, the sensitivity to pressure is relatively high, and the result can be correctly perceived. Once the degree of perception is high, it can simulate the characteristics of the input picture to the maximum. Its pressure-sensitive base has reached the 2048 level, and it is a very sensitive pressure-sensitive device [8].

The IP address refers to the Internet Protocol address. The IP address is a unified address format provided by the IP protocol. It assigns a logical address to each network and each host on the Internet to shield the differences in physical addresses.

The network teaching system is based on the standardized management of one-to-many teaching mode. The students connect with the teacher's teaching system through the network interface and the corresponding network IP, which effectively reduces the repeated teaching of a certain aspect of the knowledge generated by the teacher's teaching. In addition, this teaching system has the function of preserving teaching content. We can understand and absorb the knowledge taught by the teacher in the spare time if we do not get a full understanding in the classroom. This not only reduces the repetition of the teacher's lectures but also ensures the repetitive learning of the information [9]. Figure 2 shows the data interaction between the WEB server and the back-end database.

3.2. Digital Image Processing Technology. The processing of distorted images belongs to the category of digital image processing. Distorted images in remote teaching activities are mostly caused by low-priced and low-quality digital image acquisition equipment used by educational institutions or students due to economic conditions, such as low-pixel cameras and scanners. Especially low-cost cameras mostly use lenses with optical distortions, and they cannot get satisfactory images when collecting static images, let alone collecting dynamic images. In order to improve the accuracy of quantitative analysis such as pattern matching and image detection and to minimize the influence of image distortion on teaching interaction, it is necessary to correct this type of distortion.

The remote interactive image processing system will undergo some changes when processing information. For a one-dimensional signal $f(t)$ that changes with time, if its maximum frequency range is fM, and the sampling period is T, the ideal sampling process is described as follows:

$$f_n(t) = f(t)\cdot\sum_k \delta(t-kT). \tag{1}$$

Here, $f_n(t)$ represents the signal discrete in time after sampling.

The corresponding Fourier transform is as follows [10]:

$$F_n(f) = F(f) * \left[\frac{1}{T}\sum_k \delta\left(f-\frac{k}{T}\right)\right] = \frac{1}{T}\sum_k F\left(f-\frac{k}{T}\right). \tag{2}$$

Compared with one-dimensional sampling, the process of converting the scene captured by the camera into discrete electrical signals is actually a two-dimensional sampling process [11]:

$$\begin{aligned} s(x,y) &= \sum_m\sum_n \delta(x-m\Delta x, y-n\Delta y),\\ s(x,y) &= \sum_m\sum_n \delta\left(x-m\Delta x-\frac{1}{2}m\Delta x, y-n\Delta y\right). \end{aligned} \tag{3}$$

Sampling $f(x, y)$ to get the sampled image $f_s(x,y)$ is as follows:

$$f_s(x,y) = f(x,y)\cdot\sum_m\sum_n \delta(x-m\Delta x, y-n\Delta y). \tag{4}$$

Suppose the two-dimensional sampling function $u(x, y)$ and its Fourier transform $S(u,v)$ are as follows:

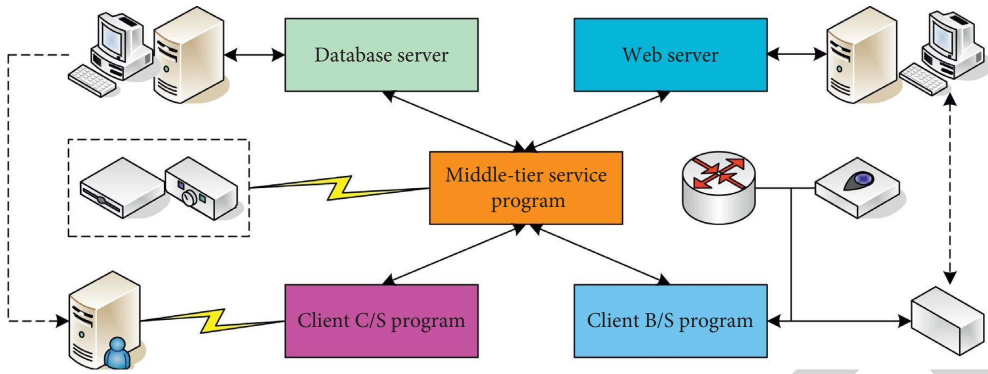


Figure 1: System structure diagram.

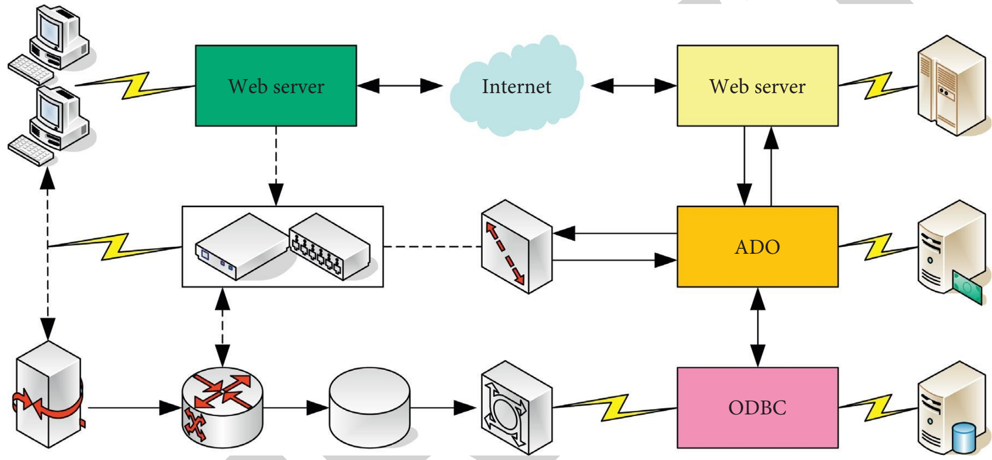


Figure 2: Data interaction between the WEB server and back-end database.

$$u(x, y) = \sum_m \sum_n \eta(x + m + \Delta x, y + n + \Delta y),$$
$$S(u, v) = \frac{1}{\Delta x}\frac{1}{\Delta y}\sum_m \sum_n \delta\left(u - m\frac{1}{\Delta x}, y - n\frac{1}{\Delta y}\right). \tag{5}$$

Assuming that the two-dimensional sampling pulse is $M \times N$ dimension, it is expressed as follows:

$$s_p(x, y) = \sum_{m=1}^{M}\sum_{n=1}^{N} p(x - m\Delta x, y - n\Delta y). \tag{6}$$

Here, $s(x, y)$ is two-dimensional Dirac sampling function [12].

The sampled image can be expressed as follows:

$$\begin{aligned} F_p(x, y) &= f(x, y) \cdot s_p(x, y) \\ &= \sum_{m=0}^{M}\sum_{n=0}^{N} f(x, y)p(x - m\Delta x, y - n\Delta y). \end{aligned} \tag{7}$$

According to the convolution theorem,

$$F_p(u, v) = \frac{1}{4\pi^2}F(u, v) * [S(u, v) \cdot P(u, v)]. \tag{8}$$

Interpolation error is as follows [13]:

$$Rn(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x - x_i)(\min(x_i) \le \xi \le \max(x_i)). \tag{9}$$

The discrete is defined as follows (the original image size is $M \times N$).

Positive transformation is as follows:

$$f(u, v) = \sum_{u=0}^{M}\sum_{v=0}^{n-1} f(v, v)e^{-2j\lambda(nx/m + vy/n)}. \tag{10}$$

The inverse transformation is as follows:

$$f(x, y) = \sum_{u=0}^{M}\sum_{v=0}^{n-1} f(x, y)e^{-j\lambda(nx/m + vy/n)}. \tag{11}$$

Assuming that the original image is F and the template is G, then

$$NC(x, y) = \frac{\sum(|F(x, y) - E(F)| \times |G(x, y) - E(G)|)}{\sqrt{\sum(F(x, y) - E(F))^2 \sum(G(x, y) - E(G))^2}}. \tag{12}$$

Take $n + 1$ reference image points $f_k(x, y)$ on a curve in any image, then the curve uses a binary n-th Lagrange interpolation polynomial, which can be expressed as follows [14]:

$$L(x, y) = \kappa \sum_{k=0}^{n} f_k(x, y) l_k(x, y). \tag{13}$$

Here, $l_k(x, y)$ is a binary n-th order interpolation shape function.

The gradient weight coefficient is defined as follows:

$$\begin{aligned}
H_L &= \frac{1}{\sqrt{1 + \alpha\left(\left|f_{i+1,j} - f_{i-1,j+1}\right| + \left|f_{i+1,j+2} - f_{i+1,j-2}\right|\right)}},\\
H_r &= \frac{1}{\sqrt{1 + \alpha\left(\left|f_{i,j} - f_{i+2,j+1}\right| + \left|f_{i+2,j+2} - f_{i+1,j-1}\right|\right)}},\\
V_u &= \frac{1}{\sqrt{1 + \alpha\left(\left|f_{i,j} - f_{i,j-1}\right| + \left|f_{i+1,j+1} - f_{i+1,j-1}\right|\right)}},\\
V_l &= \frac{1}{\sqrt{1 + \alpha\left(\left|f_{i,j+1} - f_{i+1,j}\right| + \left|f_{i+1,j+1} - f_{i+1,j+2}\right|\right)}}.
\end{aligned} \tag{14}$$

Among them, α is the sharpness parameter, which controls the sharpness of the image, and the value range is between [0, 1]. When α is 0, the gradient weight becomes 1. As α increases, the gradient weight coefficient decreases to below 1, resulting in a sharper image [15].

The bicubic interpolation function is as follows:

$$G(x, y) = \sum_{n=1}^{2} \sum_{m=-1}^{2} f_{i+m} p(s) p(t). \tag{15}$$

In the formula, $P_{m+1}(s), P_{n+1}(t)$ is the interpolation sum.

3.3. Design of Art Material Management Function

3.3.1. User Group Management Function. This system has a user group management function. A user group can have multiple users, and at the same time, a user can belong to multiple different user groups. The difference between users lies in the group of permissions. There are six levels in total. Among them, levels 1–5 are the authority within the user group, level 1 is the lowest user level, and level 6 is the user group leader. Users with level 6 can manage all user information in this group.

3.3.2. Material Management Function. This function realizes the operations on the material data table: the user can add and delete the information content of the material and can also select the classification of the material for modification. When modifying the group name and category of the material, the level must also be determined. There are 5 levels of material, divided into 1–5 levels. When the user group name is the same as the material group name, the system judges whether the material can be accessed current and level. If the user current material, the user can access it; otherwise it cannot be accessed.

3.3.3. Download Management Function. Because of the "buffer library architecture" when downloading, the file is stored in the buffer, and the downloaded material information is stored in the buffer table. After the user operation is completed, the system administrator will migrate the data into the formal table [16].

The E-R model diagram of the database conceptual structure independent of the actual data model structure must be transformed into the logical structure E-R diagram for database application design. System's appreciation resource information design is shown in Table 1.

3.4. Test Environment. Operating system: Microsoft Windows 7

Database: SQL Server 2008

Test related software: MyEclipse 10, Microsoft Office 2010, tomcat 7.0

3.5. System Test and Operation. Unit testing is mainly to test various controls, components, public functions, and reusable modules. Interface testing is mainly to understand whether the designed software interface can correctly guide the user to actively complete the operation. It needs to have a guiding role: to allow the user to work in a comfortable environment [17]. Integration testing is mainly used for the cooperation between systems and components. The security and its data cannot be stolen, whether the privacy can be protected, and whether the data can be encrypted correctly [18]. The server creation link is mainly based on the IP creation of the corresponding server, and the corresponding students are notified to go to the class, and the students add the link by themselves through the user information prompted by the teacher [19, 20]. In addition, the administrator user is responsible for teachers and students to conduct proper management for the entire user. First of all, the corresponding search can be made for the teacher users and student users of the art professional network teaching system. Secondly, the corresponding teachers and students can be appropriately added according to the needs. Once there are errors, the teachers and students can be modified and deleted appropriately according to the corresponding information.

4. Results of the Remote Interactive Art Teaching System

In the survey, it is found that the design basis of art teachers is the theoretical basis of teaching, the needs of teaching practice, the characteristics of students in each grade, and the content of teaching materials. Among them, "teaching practice needs" accounted for the highest proportion (96.2%), followed by "the characteristics of students in each grade" which accounted for 92.5% of the total number of surveys. The proportions of "based on the content of

Table 1: System appreciation resource information design.

Field name	Data type and length	Chinese meaning
ID	Int (8)	Self-incrementing ID
Type	Int (8)	Type
Title	Varchar (50)	Title
Content	Varchar (500)	Content
URL	Varchar (50)	Link

Table 2: The focus of art teachers' teaching.

Focus	Frequency	Percentage
Teaching theory basis	21	39.6
Teaching practice needs	51	96.2
Characteristics of students in each grade	49	92.5
Based on the content of the textbook	31	58.5

teaching materials" and "needs for teaching practice" are 58.5% and 39.6%, respectively. The survey results show that art teachers put a large proportion of the actual needs of teaching and the characteristics of students when designing their courses, so they ignore the support of teaching theories and teaching materials to varying degrees. The focus of art teachers' teaching is shown in Table 2.

In the survey, we can clearly and intuitively see that the number of art classes students take each week is mainly concentrated in "1–2 periods" and "3–4 periods." They each account for 35% and 58% of the total, which is more than 90% of the total (students' weekly art class 1–4 section statistics are shown in Figure 3(a)). This reflects the good start of art classes. However, the proportion of "5–7 periods" is also relatively large, indicating that some schools may have occupation and suspension of classes (students' weekly art class statistics of 5–7 sessions are shown in Figure 3(b)).

Students' favorite art classes are sketching, paper-cutting, clay sculpture, paper prints, pottery, watercolor, Chinese painting, and calligraphy. The most disliked are appreciation classes and photography. The love range of students' art projects is shown in Figure 4.

In a comparative study of male and female classifications, boys' favorite items are sketching (86.63%), copying (97.57%), painting (97.87%), clay sculpture (93.31%), pottery (84.80%), and watercolor (86.63%). Boys' least favorite item is knitting (32.83%) (boys' hobby items are shown in Figure 5(a)). Girls' favorite items are clay sculpture (88.43%), paper-cutting (86.35%), pottery (89.32%); girls' least favorite items are sketching (10.39%), copying (15.73%), appreciation class (8.01%). Girls' love for knitting and paper-cutting is significantly higher than that of boys by about 30%, while boys' love for sketching, copying, painting, and calligraphy is significantly higher than that of girls by more than 20% (girls' hobby items are shown in Figure 5(b)). It can be seen from the survey results that girls are more inclined to the more gentle course content, while boys prefer the learning content of calligraphy and painting.

The system designed in the article includes a forum module, where both student users and teacher users are regarded as registered users. The test results of the forum management function are shown in Table 3.

According to the performance design of the system, the main function module needs to define the main data table, and several auxiliary data tables need to be designed. The student information design is shown in Table 4.

The oil painting masterpiece information is shown in Table 5.

Judging from the results of the survey, most of the students participating in the survey think that using the remote interactive art teaching system is helpful to complete the art class (the statistics of great help and a little help are shown in Figure 6(a)). This also fully shows that this form has a positive impact on the learning status of students (the statistics of no help and negative help are shown in Figure 6(b)).

35% of students think it is "increasing opportunities to exchange works and ideas with teachers and classmates" (statistics for supplementing and expanding the knowledge of art disciplines and increasing opportunities for the exchange of works and ideas are shown in Figure 7(a)). There are also 22% of students think it is understanding the teacher's lesson plans (to understand the basic learning situation of the teacher's teaching plans, homework completion specifications, and to improve the enthusiasm of active learning and completion of homework, the statistics are shown in Figure 7(b)).

In the survey on the teaching function of the website, 74% of the students liked the setting of the learning resource area most, and 52% liked the "work display and evaluation function" (the statistics of learning resources and work display are shown in Figure 8(a)). Another 30% of students like the design of interactive functions, but students are generally not interested in the function of "filling in learning files" (the statistics of interactive functions and filling in learning files are shown in Figure 8(b)).

In the survey on the difficulty of using the remote interactive art teaching system to learn art, 39% of the students think it is more difficult to complete the homework, 13% of students think it is the most difficult to find learning materials to complete the homework, and 22% of students think it is more difficult to be familiar with the website interface and functions. The survey results of the difficulty of learning art are shown in Figure 9.

83% of students think that the remote interactive art teaching system makes them more like art classes, 78% of students are willing to use the remote interactive art teaching system to assist in learning in art classes in the future. Another 22% of the classmates were indifferent to which way to take an art class. The willingness of students for the remote interactive art teaching platform is shown in Figure 10.

5. Discussion

The Internet allows us to better understand society, show our individuality, participate in society extensively, and meet people's entertainment needs. Network communication expands the field and objects of our communication, changes the way of communication in the past, and enriches our life experience. In the new space of online communication, we can break through the professional restrictions.

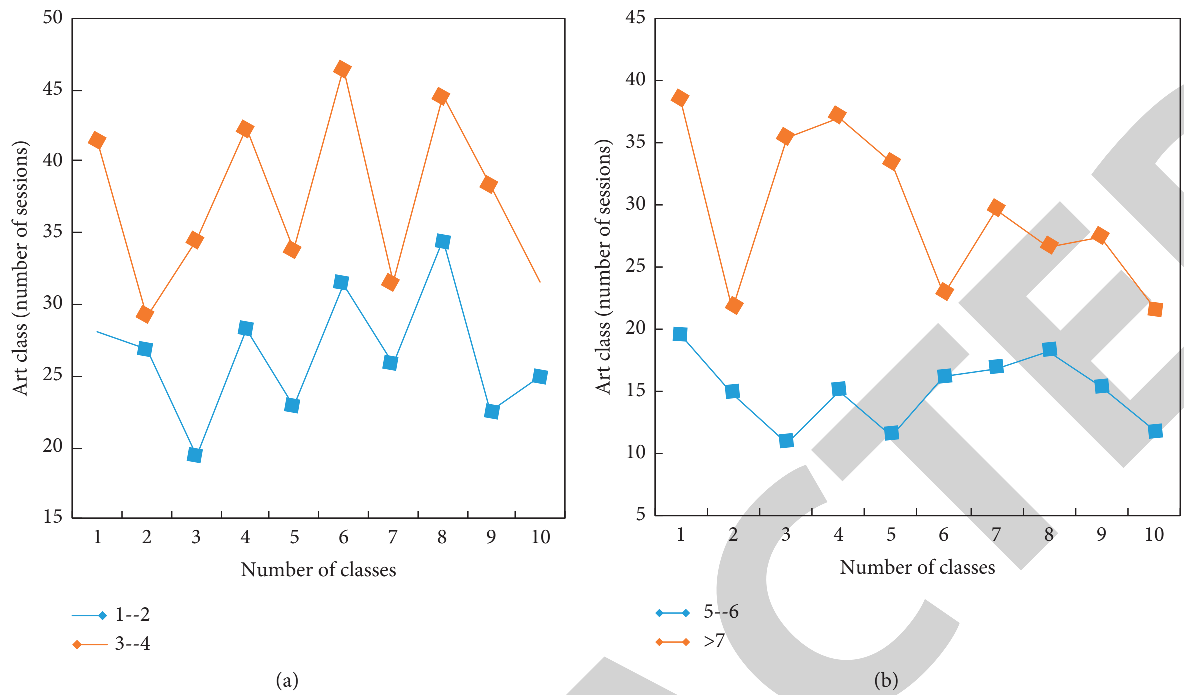


Figure 3: The number of student art classes per week. (a) Sections 1–4. (b) Sections 5–7.

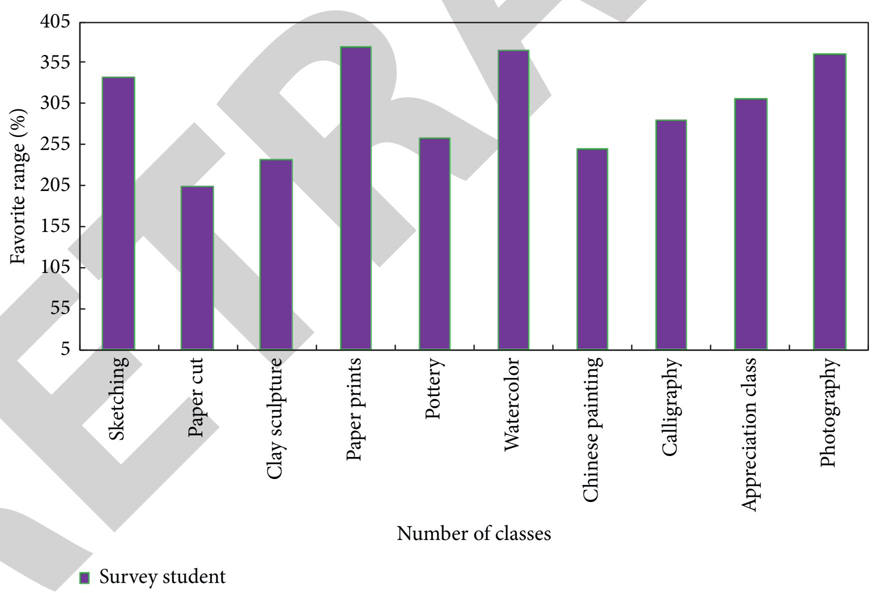


Figure 4: The range of students' passion for art projects.

As long as we want to learn, we can find learning resources and instructors at any time. The network is very useful, but of course, it is not just the above.

Under the background of the network information age, Internet technology has been fully penetrated into all walks of life, and most of the schools have introduced electronic computer technology and network information technology, relying on these two technologies to carry out curriculum teaching activities. In this way, it enriches the content of students' curriculum and greatly improves efficiency and quality, which shows that it can promote students' autonomous learning. With the improvement and prosperity of China's system reform, some teaching methods and teaching models that were not good in the past can no longer meet the current teaching needs. Therefore, many teaching methods must appear as an education model. This is a challenge that every developing country and developed country's education system must face, and it is also the historical law of the development of education systems in countries all over the world. Only by improving the country's own educational system can we have a better curriculum teaching concept. This is an

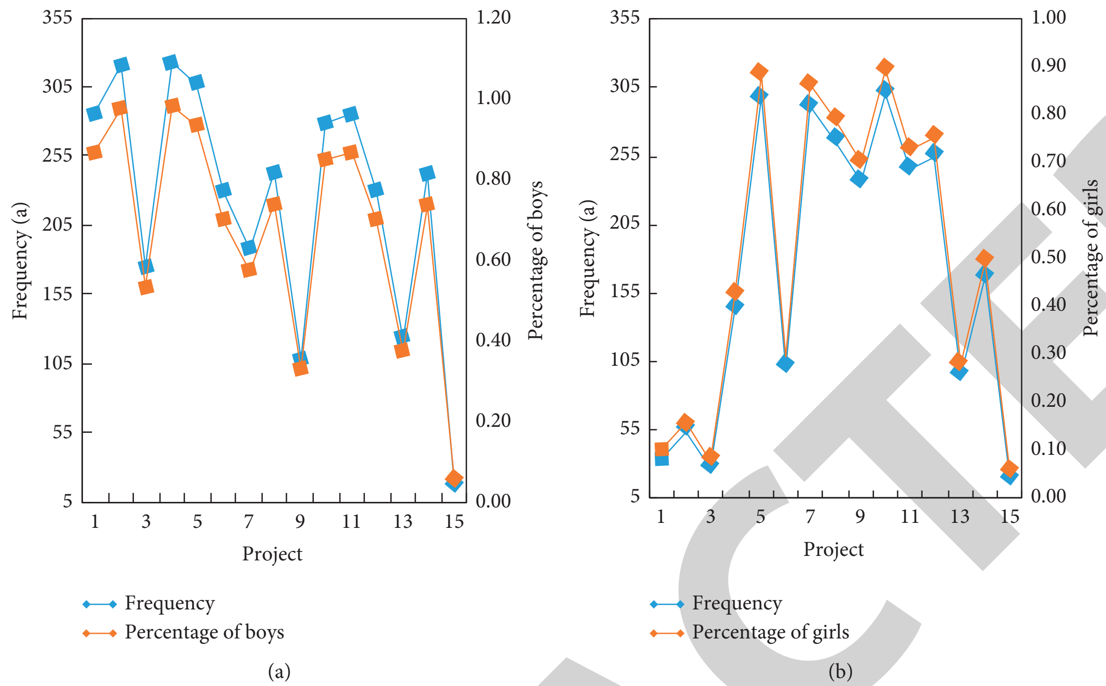


Figure 5: Comparison of hobby items. (a) Boys' hobby items. (b) Girls' hobby items.

Table 3: Test results of forum management function.

Module name	Module function	Percentage
Plate module	Browse section	92.6
	Browse the topic list	95.2
Theme module	Query subject	93.5
	Publish topic	98.5
	Edit theme	97.7
	View topic	96.6
Postmodule	Reply to post	94.8

Table 4: Student information design.

Serial number	Field name	Field type and size
1	Student ID	Number (20)
2	Name	Varchar2 (10)
3	Sex	Varchar2 (4)
4	College	Varchar2 (20)
5	Major	Varchar2 (20)
6	Class	Varchar2 (20)
7	Course	Varchar2 (100)

Table 5: Oil painting masterpiece information.

Serial number	Field name	Field type and size
1	Painting ID	Number (20)
2	Topic	Varchar2 (40)
3	Auther	Varchar2 (40)
4	Country	Varchar2 (20)

important issue that we, as today's curriculum educators, must be aware of.

This article is aimed at the network education of art colleges and universities. Compared with students of other majors, they are very special in their professionalism, teaching focus, and training of professional knowledge. The previous student tutoring methods basically adopted a one-to-one method, which not only limited the utilization of teachers' time, but also students had hired teachers to tutor themselves. Therefore, we plan to use a more intelligent way to replace the existing teaching methods. Under this condition, the teacher can arrange self-built teaching tasks according to the common learning time of everyone and timely count the students' understanding of a certain module to increase the corresponding teaching ratio to the module. In addition, we divide the courses into key points and difficult points according to the level of understanding of most students in class. This not only can bring students the maximum learning efficiency but also allows students to obtain the maximum knowledge reserve effect in the limited learning time. Therefore, we develop an intelligent network teaching system suitable for learning in art colleges.

This article is mainly based on providing a more efficient and convenient teaching system for teaching. After in-depth research and investigation and analysis of art teaching, it mainly facilitates the teacher's teaching, and secondly, saves the content taught by the teacher, so as to help the students review and find the source of the problems for the students who do not understand. In addition, the system has a focus on the teacher's lectures, assigns priorities according to the students' listening situation, briefly describes most of the problems that can be understood, and focuses on the difficult points that everyone thinks. The selection of images should be based on the characteristics and requirements of the teaching content, follow the principles of demand and applicability, and focus on being able to explain the key points and

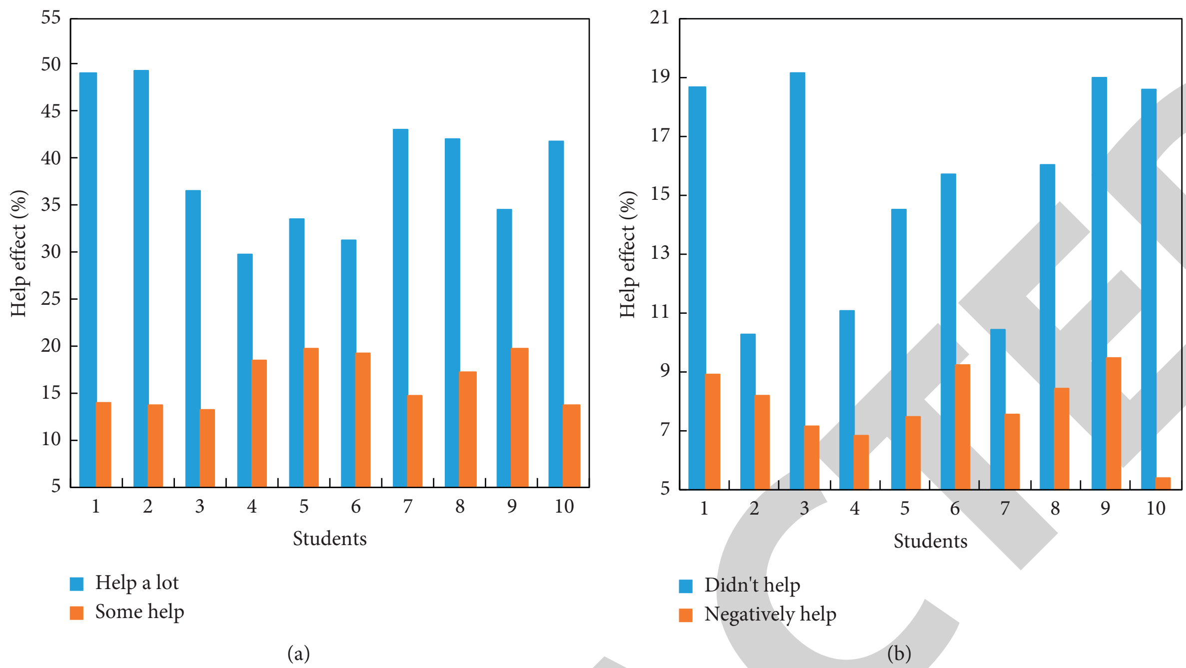


FIGURE 6: The situation of help to art class. (a) A lot of help and a little help. (b) No help and negative help.

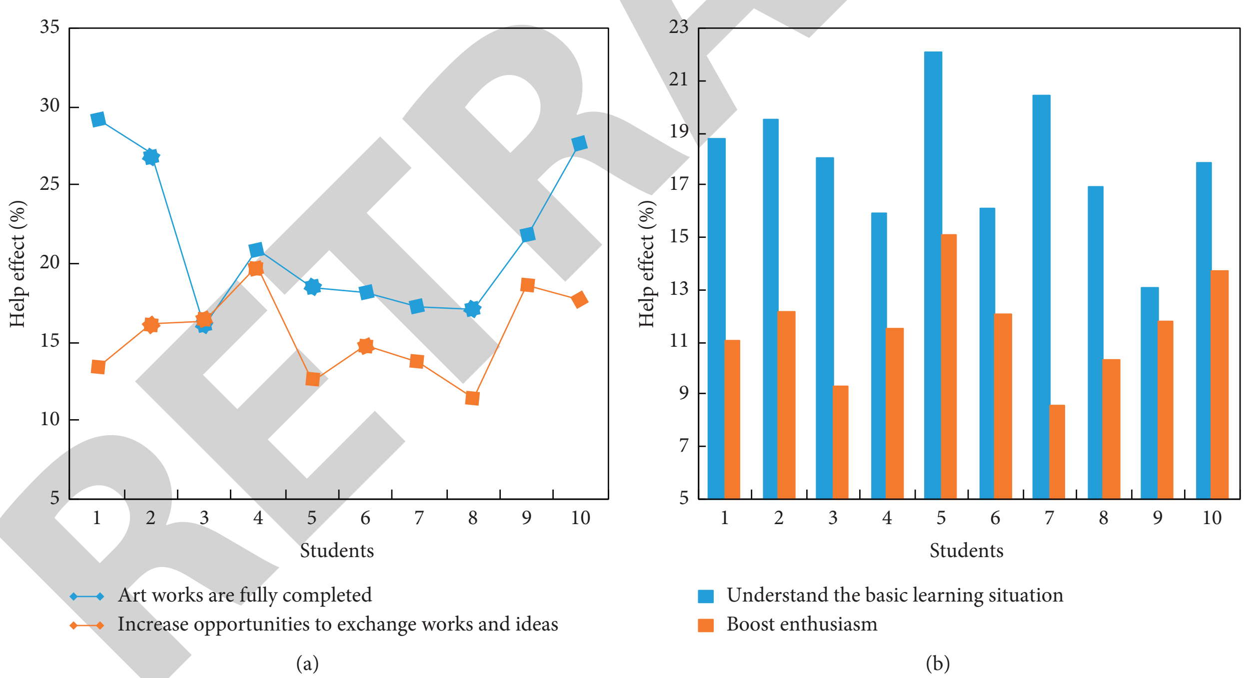


FIGURE 7: Specific help. (a) Supplement and expand the knowledge of art disciplines and increase opportunities for the exchange of works and ideas. (b) Understand the basic learning situation of the teacher's teaching plans and homework completion specifications, and improve the enthusiasm for active learning and completion of homework.

difficulties of teaching or to create a situation and stimulate interest in learning. The layout of images should pay attention to grasping the "degree," based on the principle of simplification, and use less or no images that are not related to the teaching content so as to avoid being self-defeating and causing the overflow of invalid information. More complex images should be cut off from irrelevant parts to highlight details related to the subject.

Online network teaching can not only bring students the maximum learning efficiency, but also use a more humane teaching method to learn, pay attention to every student's participation, mobilize the enthusiasm of students in

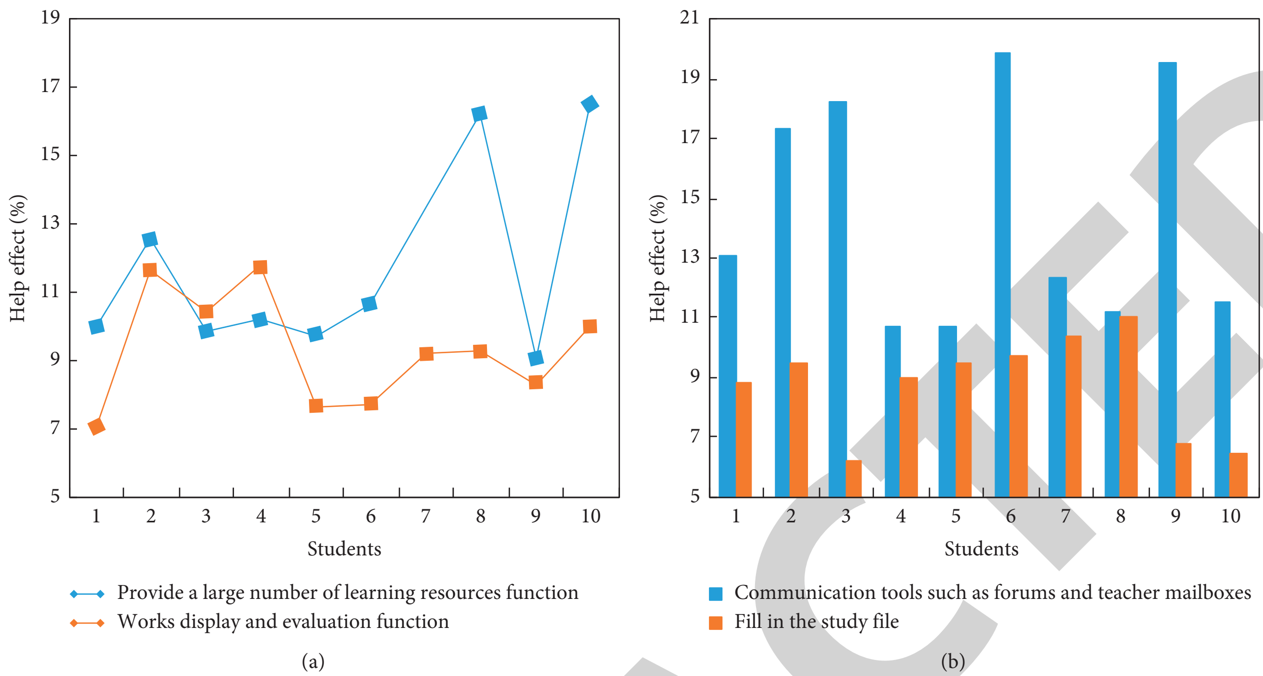


FIGURE 8: Website teaching function. (a) Statistics of learning resources and work display. (b) Interactive functions and filling in learning files.

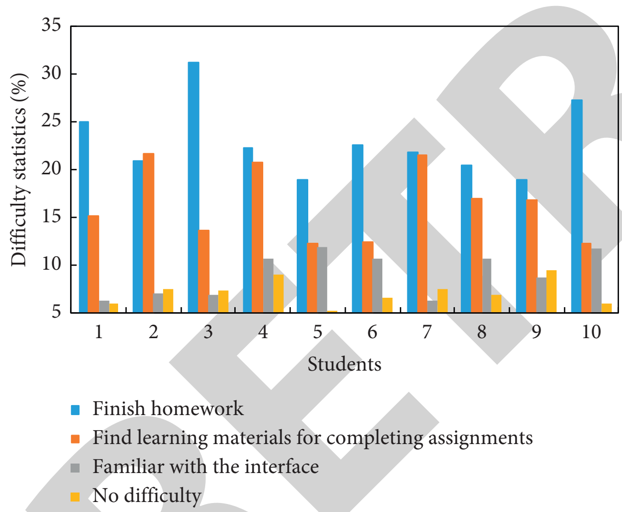


FIGURE 9: Survey results of the difficulty of learning art.

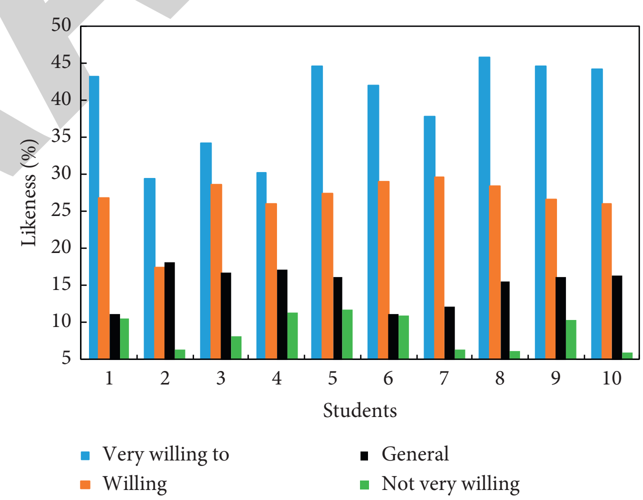


FIGURE 10: Students' willingness to remote interactive art teaching platform.

learning, and then give play to the teacher's initiative in teaching. Combining the two organically allows students to learn easily, and teachers can better grasp the key points in lectures. Therefore, the teaching methods are flexible and diverse, and a considerable effort has been made to realize the educational purpose of convenient, efficient, and happy learning. Through the network teaching system, the art is intelligent and humanized. The network teaching system uses not only advanced theoretical knowledge to transfer the traditional art teaching to the network but also uses modern technology to carry out epoch-making reforms to the traditional teaching methods. We must not only carry out the reform of art teaching but also realize the art education method based on student learning.

In the modern and contemporary world, whether in China or in other western developed countries, there are not much research on the teaching management system of art courses. So this shows that in order to increase the frequency, it is necessary to pay and study meticulously by professional teachers responsible for art course teaching management and educators from all walks of life. This may be a long, complicated, and arduous process. This process is not only an opportunity for us but also a challenge. Because any teaching method law that adapts to the educational system must withstand the practical test in the process of social development. Only by withstanding the test can we adapt to the educational trend of contributing research current social art curriculum teaching management system.

Modern distance education can serve the development of remote areas, especially the western region, and establish a policy of poverty alleviation through distance education. The Ministry of Education supports a distance education system in the western region. Modern distance education has a good foundation and conditions in this regard. Modern distance education conforms to the trend of world globalization and promotes education to the advent of the world information age. School education cannot be done behind closed doors. Modern distance education is a good way to conduct teaching and academic exchanges with foreign countries. This article is aimed at schools and researches in distance teaching, supplementing the various teaching networks currently available in schools so as to achieve the purpose of interconnection with distance teaching and distance teaching management. However, this topic is based on the actual project of a specific school as the research content. However, in order to not lose generality, various distance teaching networks currently existing in the school are discussed, and a universally applicable interconnection model and teaching management system are provided to improve the practicability and applicability of this system.

The information age calls for distance education. After years of development, schools have established various forms of auxiliary teaching facilities and methods. For example, multimedia computer network classrooms, multimedia TV teaching one-way or two-way teaching systems, campus multimedia integrated network teaching systems, stand-alone systems, campus computer network systems, and so on. All kinds of teaching systems are only used inside the school, lack of communication with the outside world, this situation of isolation from the outside world does not include information age. Education reform requires distance education, and many disadvantages of traditional education are obvious to all. In the current situation of advocating quality education, the implementation of modern distance education methods can not only overcome some of the shortcomings of traditional education but also has much practical significance.

The art discipline aesthetic education of students, cultivation of professional skills. Therefore, the demonstration of teachers between classes is indispensable. However, due to the limitation of the environment or time, the teaching of many art classes is not effective in the classroom. It is difficult to tell the whole painting process only by the teacher's dictation.

The use of technology to realize distance education has become the world's largest computer network, and tens of millions of users are enjoying the information services and resource sharing provided by the network. The agreement that constitutes one of the three elements of the network plays an important role. To realize a complete teaching function, the simple operation and convenient use of the system should also be considered. Although the campus multimedia integrated teaching network system is quite different from the traditional TV teaching network in structure, the operation, use, and management of the system are basically the same as before. Compared with the computer network, it can be said to be simple to operate, easy to manage and use, so it is convenient for ordinary schools to use.

6. Conclusion

There are two main applications of computer networks in distance education. One is to connect the teaching network with a wide area by using protocols, and the other is to use technology to build a campus network. The application of remote interactive teaching integrates various technical means. The ultimate goal so that students can learn effectively even if they are in a different place. Teacher's teaching always occupies the most important part. Long-distance interactive teaching should enable students who are both onsite and studying in different places to participate in the communication and information when teachers are teaching. This paper uses the computer software design language based on JAVAEE and Oracle database server to complete the design of the conceptual structure, the logical structure the design of the application software interface interaction based on the B/S architecture. Under the traditional education model, teachers spend a lot of time in the process of homework assignment and homework correction, the work content is cumbersome, and it is difficult to judge whether students' homework is plagiarized. Therefore, designing a set of fully functional interactive teaching systems has important practical significance for the development of modern education. The problem of each pixel in the standard image before distortion needs to be further discussed in future work.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] B. Kołodziejczak and M. Roszak, "ICT competencies for academic E-learning. Preparing students for distance education - authors' proposal," *Nephron Clinical Practice*, vol. 6, no. 3, pp. 14–25, 2017.

[2] R. Shields, "Curriculum design for distance education in the tertiary sector," *The Turkish Online Journal of Distance Education*, vol. 18, no. 2, pp. 4–15, 2017.

[3] M. Harsasi and A. Sutawijaya, "Determinants of student satisfaction in online tutorial: a study of A distance education institution," *The Turkish Online Journal of Distance Education*, vol. 19, no. 1, pp. 89–99, 2018.

[4] D. Vasilevska, B. Rivza, and R. Bogdan, "Evaluation of readiness for distance education of students in European universities," *BRAIN: Broad Research in Artificial Intelligence and Neuroscience*, vol. 8, no. 1, pp. 35–41, 2017.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9876328, 1 page
https://doi.org/10.1155/2023/9876328

Retraction

Retracted: Construction and Analysis of Urban Cultural Plane Space Mode considering Particle Swarm Cultural Scientific Computing Algorithm

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Yu and Q. Li, "Construction and Analysis of Urban Cultural Plane Space Mode considering Particle Swarm Cultural Scientific Computing Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 9838782, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9838782, 14 pages
https://doi.org/10.1155/2022/9838782

Research Article

Construction and Analysis of Urban Cultural Plane Space Mode considering Particle Swarm Cultural Scientific Computing Algorithm

Jiayong Yu[1] and Qi Li[2]

[1]*Xi'an University of Architecture and Technology, Xi'an 710055, China*
[2]*Xi'an City Planning & Design Institute, Xi'an 710082, China*

Correspondence should be addressed to Jiayong Yu; jiayongyu@xauat.edu.cn

Received 3 January 2022; Revised 10 February 2022; Accepted 16 February 2022; Published 14 March 2022

Academic Editor: Chin-Ling Chen

In the tide of economic and social internationalization and high-speed urbanization, urban culture has increasingly attracted widespread attention from all walks of life and has a very critical strategic significance in the construction of urban space. This paper aims to construct and analyze the plane space mode of urban culture based on the particle swarm cultural scientific computing algorithm. This article first explains the concept of urban cultural space. Urban cultural space uses urban spatial structure as a media carrier and uses urban cultural time for vertical extension and development; then we proposed the particle swarm cultural scientific calculation algorithm and gave the particle swarm algorithm flowchart; then based on the particle swarm optimization algorithm, the evaluation of the configuration performance of urban cultural facilities is researched and discussed, and at the same time, the evolution law of urban spatial morphology is explored based on the particle swarm optimization algorithm. The particle swarm algorithm is a random search algorithm based on group cooperation developed by simulating the foraging behavior of birds. Urban cultural space is the development of urban space based on urban culture. It plays a special and important role in urban cultural development and urban space planning. According to the statistics of the survey and research results, the utilization rate of city-level cultural facilities in City A is 77, the utilization rate of district-level cultural facilities is 72, the utilization rate of street-community cultural facilities is 69, and the overall evaluation score is 70. It shows that there are significant differences in the actual use of cultural facilities, so it is particularly important to eliminate the differences in the use of cultural facilities between urban and rural areas. In the exploration of the law of urban spatial morphology evolution, it is found that the number of college students, total real estate investment, urban population, and total commercial housing sales have a significant impact on urban spatial expansion and evolution.

1. Introduction

The strong transmission of world history and culture has brought a great impact on cultural industries all over the world, and consumer art culture and entertainment art culture have gradually penetrated into society. The existence of ethnic minorities and regional economic status and social culture of various countries have been severely affected, and the cultural traditions of ethnic minorities of various countries have also been threatened. And various art and cultural spaces are a kind of container to deal with this crisis. At the beginning of the economic era of the last century, the functions of cultural elements have gradually become clear, and various excellent cultural innovation products have gradually appeared. They are regarded as the cultural attraction of one of the three major soft powers of society, which has created benefits for the entire economic and social development. A new type of social structure was created, namely, cultural development space design, cultural entertainment space design, cultural consumption space design, and social and cultural gathering areas. How to organize and integrate this structure to promote the common development of society is a brand-new topic that must be faced in particular.

Observing the problems of cultural space planning in today's social cities, some have separated the urban cultural space planning from the overall urban space planning and separated it from regional development; there are also cities that overconstruct and develop culture, so that they ignore the urban living space and other important consumption spaces. Therefore, it is self-evident to analyze and study the importance of urban cultural space. Only by comprehensively analyzing the social conditions of urbanization development and attaching importance to the function of urban cultural construction can we propose a rational strategy for urban cultural space planning and at the same time, it is the rationality of the two to look at its positive and negative aspects. This paper first expounds the concept of urban cultural space, then proposes a particle swarm cultural scientific computing algorithm, and gives a flowchart of the particle swarm algorithm. Finally, based on the particle swarm optimization algorithm, the evaluation of the configuration performance of urban cultural facilities is studied and discussed, and the evolution law of urban space morphology based on the particle swarm optimization algorithm is also explored. This paper aims to construct and analyze the plane space mode of urban culture based on the particle swarm cultural scientific calculation algorithm, hoping to provide a practical effect on the promotion of urban culture.

According to the research progress at home and abroad, different scholars also have a certain degree of cooperation in the construction of particle swarm cultural scientific calculation algorithm and the construction of urban cultural plane space mode: Yang H proposed a new cooperative control method based on loop tracking algorithm and fuzzy control idea to realize the reconstruction of multiagent formation in 3D space. His idea is to use nonlinear loop tracking control in the longitudinal movement of the formation and use a piecewise proportional derivative (PD) controller based on fuzzy control for the cooperative control of the normal movement. The results show that this method improves the performance and accuracy of formation reconfiguration control, avoids collisions between members, and enhances the stability of the system [1]. The methodology used by M'Hammedi is an evaluation of the main methods used in the creation of the bare new city of Rabat. He experimented with urban landscape based on spatial functionalism, landscape treatment, urban composition, and architectural planning, emphasizing architectural design influenced by different forms of decorative art [2]. The purpose of Kherbouche and Djedid's research is to test the ever-changing cognition of the city's image, evaluate its persistence, and determine its relationship with the development of sustainable cultural tourism. It is tested by empirical research combining qualitative and quantitative methods. The qualitative research is based on semidirect interviews, and the quantitative research is based on statistics from the Wilaya Tourism Bureau. The results of this study show that residents' perception of the city's image is not static. It follows the same evolution process as tourists' perception of the image. To achieve the sustainable development of cultural tourism, the sustainability of the image must be verified internally and externally [3]. The research of Amen MA aims to find the internal symbol system in the urban space organization, explore the hidden space organization, and find the inner power as the aesthetic symbol of the urban space organization. The results show that, without considering the existing sign system, the results of introducing new concepts into urban spatial organization are different. This difference may be the source of misunderstanding and confusion in the spatial organization system, which needs to be understood and reorganized through the development of procedures [4]. Lee S research analyzed the urban microspatial structure by considering the complex use of the interior of the building. The significance of the research lies in taking into account the complex uses of buildings, constructing a more refined and microurban spatial structure, clarifying the structural differences in retail office land prices and price formation factors; through the analysis of the hierarchical linear model, it is found that the land price formation structure of commercial real estate may vary depending on the use of the region [5]. Naseri and Jafari Navimipour proposed a new hybrid approach to achieve efficient service composition in cloud computing. Agent-based methods are also used to combine services by identifying QoS parameters and use particle swarm optimization (PSO) algorithms to select the best services based on fitness functions. The simulation results show the performance of this method in reducing combined resources and waiting time [6]. Wang introduced the origin and background of PSO and conducted a theoretical analysis of PSO. Then he analyzed its research and application status in algorithm structure, parameter selection, topology structure, discrete PSO algorithm, and parallel PSO algorithm, multiobjective optimization PSO, and its engineering application. Finally, the existing problems are analyzed, and future research directions are proposed [7]. Wu and Song first introduced the development history and related research work of swarm intelligence algorithms at home and abroad. Secondly, the importance of particle swarm optimization algorithm in power system network reconstruction is proposed, and the basic principle, essential characteristics, and basic model of particle swarm optimization algorithm are expounded. The improved particle swarm optimization algorithm proposed by Wu and Song provides a good solution for power system network reconstruction and has important research significance for the subsequent optimization of power system and the improvement of particle swarm optimization algorithm [8]. However, these scholars did not combine the particle swarm cultural scientific calculation algorithm with the construction of urban cultural plane space mode to explain the problem but unilaterally explored their meaning. The purpose of this study is to make up for the research gap in the construction of urban cultural plane space model, in order to provide practical effects on the promotion of urban culture.

The innovations of this article are mainly reflected in the following: (1) firstly, it explained the concept of urban cultural space. Urban cultural space uses urban spatial structure as a media carrier and uses urban cultural time for vertical extension and development; (2) it proposed the particle swarm cultural scientific computing algorithm and gave the particle swarm algorithm flowchart; (3) at the same time, based on the particle swarm optimization algorithm, the evaluation of the configuration performance of urban cultural facilities is researched and discussed, and at the same time, based on the particle swarm optimization algorithm, the evolution of urban spatial morphology is explored.

2. The Construction and Analysis Method of Urban Cultural Plane Space Mode Taking Into Account the Particle Swarm Cultural Scientific Computing Algorithm

2.1. Urban Cultural Space. City cultural space is the development of urban space based on urban culture. It plays a special and important role in urban cultural development and urban space planning [8]. The formation of urban cultural space is an organic synthesis of urban development, cultural development, and industrial development. The implementation of urban cultural space development strategy is an important path for urban cultural development and an effective way for urban economic development. It is also a trend and successful experience of urban cultural development and urban space planning at home and abroad [9]. A city's cultural image is the appearance of a city's culture. It is a city form and feature that can inspire people's thoughts and emotions. It is the specific perception, overall view, and comprehensive evaluation of the city's internal and external public on the city's inherent strength, apparent vitality, and development prospects. In the process of its formation and development, the city has created its own urban culture, becoming an important part of human culture and the most active, brilliant, creative element and the crystallization of wisdom. It is the high-level expression or highest form of culture created by mankind on the Earth. The development of urban culture, especially the protection and utilization of historical and cultural heritage, the protection and utilization of scenic spots, urban ecological environment protection, urban cultural construction and activities, etc., has made great progress and development.

Regarding the definition of urban social and cultural public space, people first proposed that the city is a cultural place that occupies the corresponding material public space and at the same time has been widely recognized by the citizens and reflects the characteristics of urban public civilization [10]. It is formed by organically fusing the three basic elements of human beings as the main body of the spatial structure of social, historical, and cultural activities, activities that constitute nodes of time, and cultural places of spatial structure nodes, interacting and coexisting. It is divided according to the differentiation of the spatial structure of people's concentrated living, the difference in the intensity of collective social, historical, and cultural activities, and the spatial structure of locations. The spatial structure of urban social, historical, and cultural activities is divided into the overall urban social, historical, and cultural image space, the large-scale historical and cultural divisions of the urban internal structure, the microsized cultural areas in the urban internal structure, and the microsized cultural facilities of the urban internal structure. In terms of demand, the design of urban cultural space is divided into basic cultural space design that allows residents to survive normally, allows the city to run smoothly, and highlights the background of urban civilization; it is known as the iconic cultural space of the iconic blocks, facilities, and venues of the city's "image project" [11]. The spatial structure of urban culture is also regarded as the cultural spiritual dimension of urban spatial structure, the highest form of urban spatial structure, and an important way of expressing the interaction between material and spiritual activities of urban people. It is mainly composed of three dimensions: the rich historical and cultural spatial structure of urban traditional historical culture, the diverse real spatial structure of urban reality and culture, and the extended spatial structure of urban social culture in the future [12]. At the same time, the urban cultural space is regarded as the existence form expressed by the material-social cultural form created by people, and it has the six characteristics of location, economy, humanity, symbolism, citizenship, and modernity [13]. Figure 1 shows the urban cultural landscape.

It can be seen that the urban cultural space takes the urban space structure as the medium carrier and uses urban cultural time to carry out vertical extension and development [14]. According to the broad and narrow categories of urban culture, we can define the category of urban cultural space as broad urban cultural space and narrow urban cultural space [15]. That is, the broad sense of urban cultural space refers to the composition and combination of living space constituted by the life energy and spiritual wealth provided by the urban subject in the historical development of the city, while the narrow urban cultural space refers to the long time in the city by the urban subject. The historical development of the city has cultivated and produced a social and cultural space organization composed of unique urban common thinking, values, basic beliefs, urban spirit, behavioral norms, and other valuable spiritual wealth [16].

2.2. Particle Swarm Cultural Scientific Computing Algorithm. This paper aims to construct and analyze the plane space pattern of urban culture based on particle swarm algorithm. As a system composed of various types and levels, the urban cultural facility system should first consider the integrity of its functional structure in each area of the city when evaluating the fairness of its facilities. In order to judge whether it can meet the multilevel cultural needs of urban residents, based on the particle swarm algorithm, the cultural space users are simulated as a flock of birds, the urban cultural space is simulated as food, and the behavior of searching for cultural facilities is simulated as birds searching for food.

The particle swarm optimization algorithm is an intelligent algorithm based on an iterative model. Its basic idea is

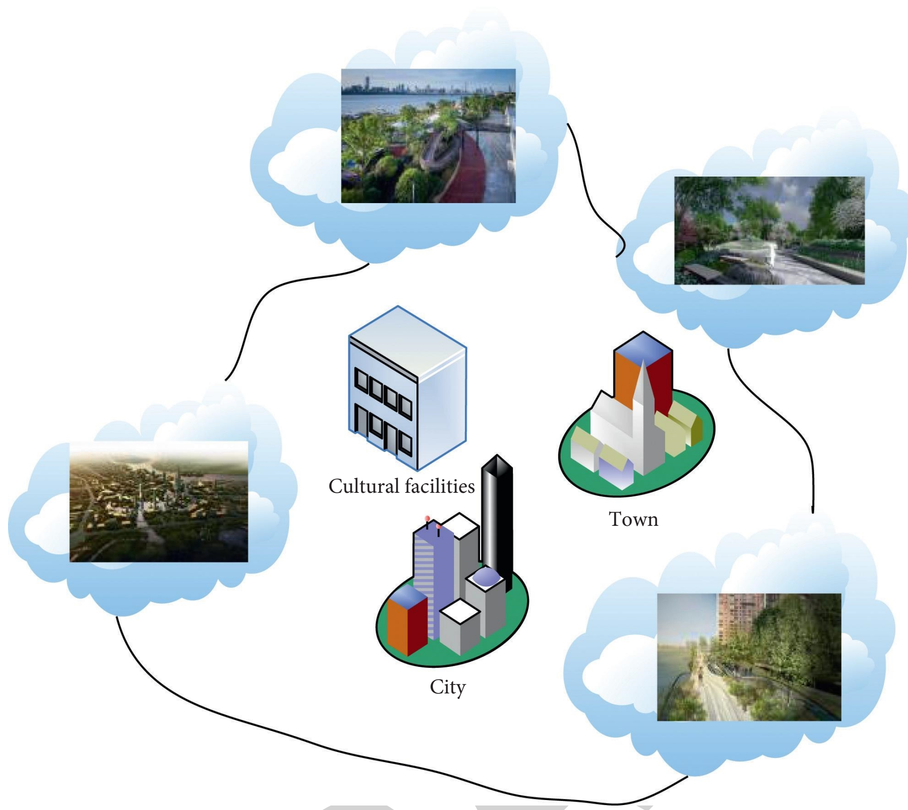


Figure 1: Urban cultural landscape.

to randomly initialize each particle that has memory but lacks volume and mass. Each particle represents a possible solution in the swarm optimization problem, and a fitness distribution value is defined for it by the fitness function. The fineness of the particles is also judged according to the size of the fitness distribution value [17]. Each particle has the ability to remember and can adjust its trajectory according to its current position, information sharing between companions, and the best position it has experienced in memory. After iteration, it keeps getting closer to the best position and finally reaches the best position [18]. The so-called optimal position is a point in the solution space that corresponds to the smallest or largest value of the fitness function. Figure 2 shows the energy-efficient power distribution method based on particle swarm algorithm.

Among them, genetic algorithm is a computational model that simulates the biological evolution process of natural selection and genetic mechanism of Darwin's biological evolution theory and is a method to search for the optimal solution by simulating the natural evolution process. Among them, the shortcomings of genetic algorithm include coding irregularities and inaccuracy of the representation of the code: a single genetic algorithm code cannot fully express the constraints of the optimization problem; being easy to prematurely converge, the accuracy and feasibility of the algorithm, computational complexity, etc.: there is no effective quantitative analysis method yet. Among them, the advantage of evolutionary computing of particle swarm optimization algorithm is that it can deal with some problems that traditional methods cannot handle. For PSO to simulate the predation behavior of a flock of birds, it means that a flock of birds is searching for food at random, and there is only one piece of food in this area. All the birds do not know where the food is. But they know how far they are from the food. So what is the optimal strategy for finding food? The easiest and most effective way is to search the area around the bird closest to the food.

For example, nondifferentiable node transfer function or no gradient information exists. But the disadvantages are as follows: the performance is not particularly good on some problems; the coding of network weights and the selection of genetic operators are sometimes troublesome.

The mathematical description of the basic particle swarm algorithm is as follows:

Assuming that the population size of the particle is M, the position information of the nth particle in the D-dimensional space is used to represent a_n:

$$a_n = (a_{n1}, a_{n2}, \ldots, a_{nd}, \ldots, a_{nD}). \tag{1}$$

The velocity of the nth particle is used to represent k_n:

$$k_n = (k_{n1}, k_{n2}, \ldots, k_{nd}, \ldots, k_{nD}). \tag{2}$$

Therefore, at the time $h+1$, the flight speed of the nth particle in the d-dimensional subspace k_{nD} is adjusted according to the following formula:

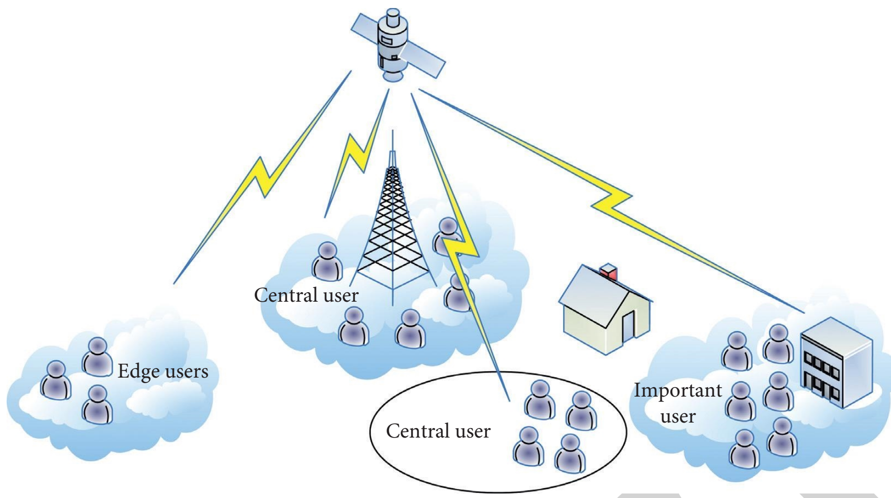


Figure 2: Energy-efficient power allocation method based on particle swarm algorithm.

$$k_{nD}(h+1) = k_{nD}(h) + e_1 w_1(h)\left(Q_{nd}(h) - a_{nD}(h)\right) + e_2 w_2(h)\left(Q_{jd}(h) - a_{nD}(h)\right). \tag{3}$$

$$\begin{cases} k_{nD}(h+1) = k_{\max}, & \text{if } k_{nD}(h+1) > k_{\max}, \\ k_{nD}(h+1) = -k_{\max}, & \text{if } k_{nD}(h+1) < -k_{\max}. \end{cases} \tag{4}$$

w_1, w_2 are a random number distributed between [0, 1].

During the search process, the particles integrate their own previous flight experience and the experience of their companions. Finally, determine your own flight speed according to formula [19]. Figure 3 is a schematic diagram of particle position update. The particle position update formula is as follows:

$$a_{\mathrm{nd}}(h+1) = a_{\mathrm{nd}}(h) + k_{\mathrm{nd}}(h). \tag{5}$$

The particle motion is coordinated by (3)–(5), and the optimal solution of various optimization problems is finally obtained by the iterative mode [20].

From the velocity update formula and position update formula, it can be seen that each dimension in the search range is independent of each other, so the convergence analysis of the algorithm can be simplified to one dimension [21]. And supposing that, in the population except for the nth one, the rest of the individuals remain motionless, then the behavior of a single individual can be studied, so the subscript n can be omitted. In order to simplify the calculation, suppose that the historical optimal position of the individual itself and the historical optimal position of the group remain unchanged, denoted as q_s and j_s; let $\mu_1 = e_1 w_1$, $\mu_2 = e_2 w_2$. Then formula (3) and formula (5) can be simplified as

$$\begin{aligned} k(h+1) &= lk(h) + \mu_1\left(Q_s - a(h)\right) + \mu_2\left(Q_s - a(h)\right), \\ a(h+1) &= a(h) + k(h+1). \end{aligned} \tag{6}$$

It can be obtained from the above formula:

$$\begin{aligned} k(h+2) &= lk(h+1) + \mu_1\left(Q_s - a(h+1)\right) \\ &\quad + \mu_2\left(Q_s - a(h+1)\right), \\ a(h+2) &= a(h+1) + k(h+2), \end{aligned} \tag{7}$$

Substituting (7) and (8) into (9) we get

$$\begin{aligned} a(h+2) &= a(h+1) + k(h+2) \\ &= a(h+1) + \mathrm{lk}(h+1) + \mu_1\left(Q_s - a(h+1)\right) \\ &\quad + \mu_2\left(Q_s - a(h+1)\right) \\ &= a(h+1) + l\left(a(h+1) - a(h)\right) \\ &\quad + \mu_1\left(Q_s - a(h+1)\right) + \mu_2\left(Q_s - a(h+1)\right) \\ &= (l - \mu_1 - \mu_2 + 1)a(h+1) - \mathrm{la}(h) \\ &\quad + \mu_1 Q_s + \mu_2 Q_s. \end{aligned} \tag{8}$$

Converting formula (10) into a standard form is

$$a(h+2) + (l - \mu_1 - \mu_2 + 1)a(h+1) + la(h) = \mu_1 Q_s + \mu_2 Q. \tag{9}$$

The above formula is a second-order nonhomogeneous difference equation with constant coefficients, so its characteristic equation can be analyzed [22].

Solve the characteristic equation of (9): $\gamma^2 + (-l + \mu_1 + \mu_2 - 1)\gamma + l = 0$. According to the solution of the quadratic equation in one variable, we can get

$$\begin{aligned} \gamma_1 &= \frac{l - \mu_1 - \mu_2 + 1 + 1\sqrt{\Delta}}{2}, \\ \gamma_1 &= \frac{l - \mu_1 - \mu_2 + 1 - 1\sqrt{\Delta}}{2}. \end{aligned} \tag{10}$$

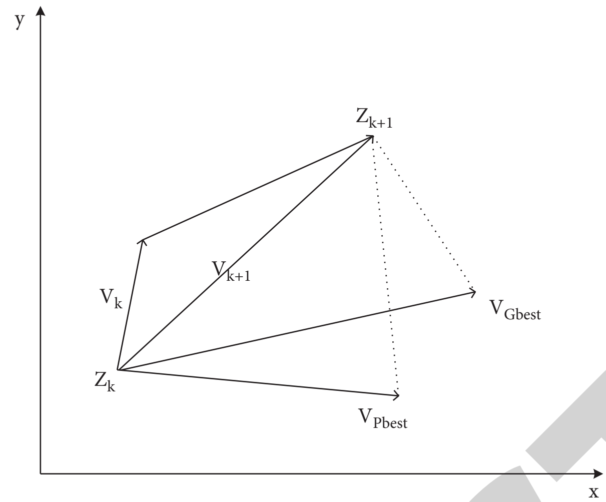


Figure 3: Schematic diagram of particle position update.

Among them $\Delta = (-l + \mu_1 + \mu_2 - 1)^2 - 4l$; if $\Delta \geq 0$, then γ_1 and γ_2 are real numbers; if $\Delta \leq 0$, then γ_1 and γ_2 are complex numbers. At this time formula (9) can be written as

$$a(h) = m_0 + m_1\gamma_1^h + m_2\gamma_2^h. \tag{11}$$

The same can be obtained:

$$k(h) = n_1\gamma_1^h + n_2\gamma_2^h. \tag{12}$$

For the convenience of description, if γ_1 and γ_2 are real numbers, $|\gamma_1|$ and $|\gamma_2|$ represent the absolute values of γ_1 and γ_2, respectively; if γ_1 and γ_2 are complex numbers, $|\gamma_1|$ and $|\gamma_2|$ represent the modulus of γ_1 and γ_2, respectively.

If and only if

$$\max\left(\left|\gamma_1\right|, \left|\gamma_2\right|\right) \leq 1, \tag{13}$$

the particle swarm optimization algorithm converges.

Proof. Find the limit of the following equation:

$$\begin{aligned}\lim_{h\longrightarrow\infty} a(h) = \lim_{h\longrightarrow\infty}\left(m_0 + m_1\gamma_1^h + m_2\gamma_2^h\right) = m_0 \\ + m_1 \lim_{h\longrightarrow\infty}\gamma_1^h + m_2 \lim_{h\longrightarrow\infty}\gamma_2^h.\end{aligned} \tag{14}$$

When $\max(|\gamma_1|, |\gamma_2|) > 1$, the limit value of (13) does not exist, so the trajectory of the particle diverges.

When $\max(|\gamma_1|, |\gamma_2|) = 1$, we can get $\lim_{h\longrightarrow\infty} a(h) = m_0 + m_1 + m_2$, or $\lim_{h\longrightarrow\infty} a(h) = m_0 + m_1$, or $\lim_{h\longrightarrow\infty} a(h) = m_0 + m_2$, so the trajectory of the particle converges.

When $\max(|\gamma_1|, |\gamma_2|) < 1$, we get $\lim_{h\longrightarrow\infty} a(h) = m_0$, and the trajectory of the particle also converges [23].

In the particle swarm optimization algorithm, the position of the particle represents the solution searched by the algorithm in the search space, so the proposition is established.

The convergence area is

$$\begin{aligned} &l < 1, \mu_1 + \mu_2 > 0, \\ &2l - \mu_1 - \mu_2 + 2 > 0. \end{aligned} \tag{15}$$

The above convergence range is also the main parameter selection range of the algorithm [24]. The key issue that affects the calculation performance and effect is the selection of parameters. The above analysis methods of convergence provide a basis for the selection of parameters [25]. The position of the current optimal predicted solution obtained by the particle itself and the position of the current optimal predicted solution of the entire population are dynamically changing. Although there are differences between the two, the selection of parameters can still be guided in the convergence area.

If the particle swarm optimization algorithm converges, the particle speed will either gradually decrease from the initial value to 0 or iterate at the initial speed until the end of the algorithm.

Proof. Find the limit of formula (15):

$$\begin{aligned}\lim_{h\longrightarrow\infty} k(h) = \lim_{h\longrightarrow\infty} n_1\gamma_1^h + n_2\gamma_2^h = n_1 \lim_{h\longrightarrow\infty}\gamma_1^h \\ + n_2 \lim_{h\longrightarrow\infty}\gamma_2^h.\end{aligned} \tag{16}$$

If the algorithm converges, then

$$\max\left(\left|\gamma_1\right|, \left|\gamma_2\right|\right) \leq 1. \tag{17}$$

When $\max(|\gamma_1|, |\gamma_2|) < 1$, the limit value of formula (13) is 0; that is, the particle velocity gradually decreases from the initial value to 0.

When $\max(|\gamma_1|, |\gamma_2|) < 1$ is used, $\lim_{h\longrightarrow\infty} k(h) = n_1 + n_2$, $\lim_{h\longrightarrow\infty} k(h) = n_1$, or $\lim_{h\longrightarrow\infty} k(h) = n_2$ can be obtained. From the initialization condition of (12), it can be known that the particle has been iterated at the initialization speed until the end of the algorithm.

If an individual's current position, the individual's historical best predicted position, and the group's historical best predicted position are all the same, the individual will leave the best historical area because its previous speed and inertia weight are not zero. So the algorithm does not converge; if the individual's previous velocity is very close to zero, once the number of particles catches up with the highest particle of the current population, the diversity of the population will gradually decrease, all particles will be concentrated in the same place and stagnant, and the overall optimization process will appear. The situation is stopped, and the global optimal predictive value cannot be found. In this case, most of the premature phenomenon will be caused; and if the individual speed iterates at the initialization rate until the end of the particle swarm optimization algorithm, it is equivalent to no longer affecting the individual's own cognitive part and the social part. It is not conducive to the search of the global optimal prediction value, and the adaptability of the algorithm will also be greatly reduced [26].

Based on the above analysis, it can be known that the particle swarm optimization algorithm cannot guarantee the convergence to the global optimal value. It can only show that the algorithm will eventually converge to the current population optimal solution; that is, if the algorithm does not search for the global optimum before converging, premature convergence will occur. Figure 4 shows the flowchart of the particle swarm algorithm.

3. The Construction of the Urban Cultural Plane Space Model and the Analysis of the Experimental Results Taking Into Account the Particle Swarm Cultural Scientific Computing Algorithm

3.1. Evaluation of the Performance of Urban Cultural Facilities Allocation Based on Particle Swarm Optimization Algorithm. This paper aims to construct and analyze the plane space pattern of urban culture based on the particle swarm cultural scientific computing algorithm. Particle swarm intelligence algorithm generally has good versatility. Due to the variety of practical application problems, the mathematical properties of many of them are difficult to determine. The optimization process of particle swarm intelligence algorithm does not depend on the strict mathematical properties of the problem itself, such as continuity and derivability. It also does not need precise mathematical description of objective function and constraints, which makes the particle swarm intelligence algorithm have better generality.

The urban cultural facility system is a system composed of multiple types and multiple levels. The evaluation of the fairness of its facilities must first consider the integrity of its functional structure in each area of the city. It is used to judge whether it can meet the multilevel cultural needs of urban residents. The functional settings of cultural facilities mainly include reading and reading, cultural performances, knowledge lectures, cultural exhibitions, leisure and entertainment, and cultivating sentiments. Urban residents of different backgrounds have different preferences for cultural functions. The evaluation of the rationality of cultural facilities functions is mainly judged from the perspective of residents' needs.

This paper takes the results of the questionnaire as the basis for the evaluation of the index and divides the evaluation of the functional rationality of cultural facilities in the questionnaire into five levels: very reasonable, relatively reasonable, general, unreasonable, and very unreasonable. Figure 5 is a statistical chart of the rationality evaluation of cultural facilities in City A.

The quality of cultural facilities mainly refers to the quality of physical buildings. Evaluation factors include building appearance, internal environment, lighting and ventilation, and fire safety. The evaluation of this indicator is based on the perception of users of cultural facilities. In the questionnaire, five levels of quality, good, good, fair, poor, and very poor are set. Figure 6 shows the physical quality evaluation of cultural facilities in City A.

Through field investigations, it was found that the physical quality of large-scale cultural facilities in City A is better, and the environment comfort is higher. Among them, the newly built venues such as the City A Library, the Cathay Pacific Art Center, and the Natural History Museum have a grand appearance, spacious and comfortable interior space, full lighting, ventilation, and lighting facilities, and the fire exits comply with the architectural markings and are highly safe; although the facades of old venues such as Children's Palace and the Mass Art Center are somewhat obsolete, the internal use space is repaired in real time and can perform better functions. Figure 7 shows the physical map of cultural facilities in City A.

The quality of district-level cultural facilities is quite different, and the construction of district-level cultural centers and libraries has been paid more attention. In addition, the construction of the district-level elderly activity center is relatively lagging, and the update status is not good, resulting in the breakdown of the building entity and the problems of hardware facilities such as air conditioning and lighting. Among the basic cultural facilities, the construction quality of the sublibraries of the district-level cultural facilities belonging to the street level is better. The flow of people is attracted by the newly built cultural facilities with high environmental quality, and the dilapidated old community has become a settlement for the elderly with low economic conditions and low education level. This has led to the decline of cultural spaces in traditional communities.

Environmental comfort mainly describes the creation of cultural atmosphere. It evaluates the quality of cultural facilities based on higher-level human psychological cognition to measure whether the activity space, environment construction, and software equipment of cultural facilities can be used comfortably by residents. The evaluation of this indicator is based on the experience of users of cultural facilities. Five levels are set in the questionnaire: very comfortable, relatively comfortable, average, not very comfortable, and very uncomfortable. Corresponding the evaluation results to the scores of each level in the questionnaire, calculate the quantifiable scores. Figure 8 shows a

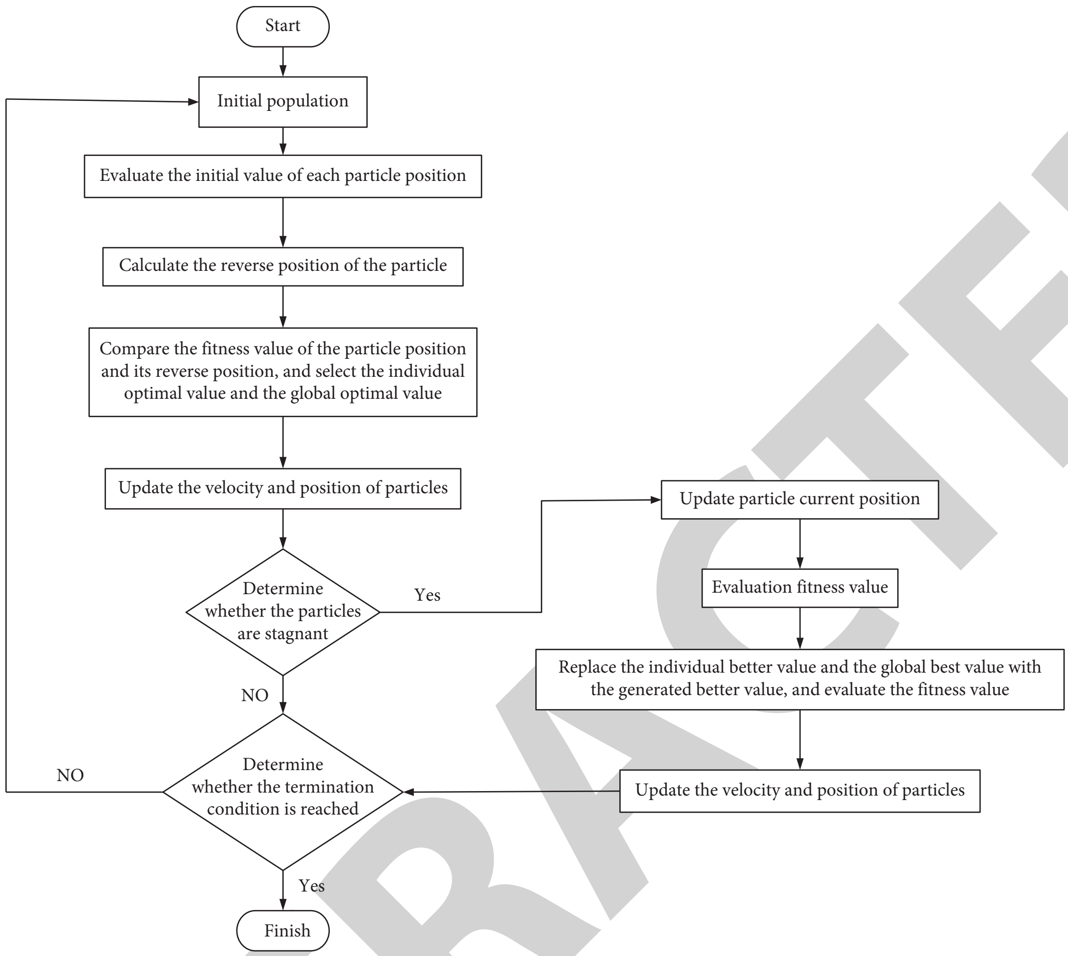


Figure 4: Flowchart of particle swarm algorithm.

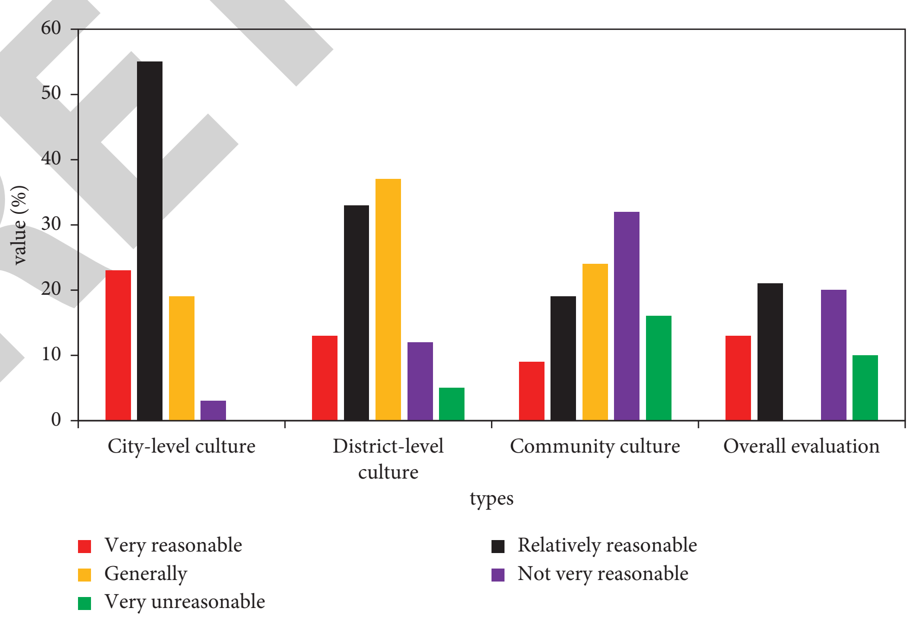


Figure 5: Evaluation of functional rationality of cultural facilities in City A.

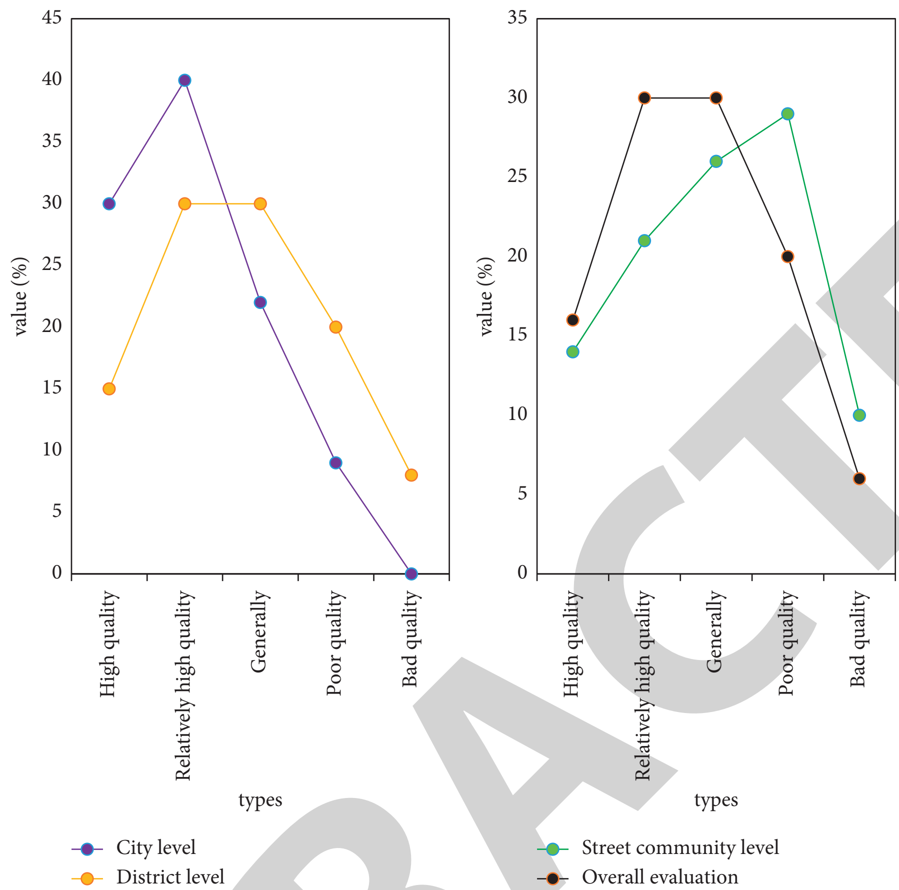


FIGURE 6: Physical quality evaluation of cultural facilities in City A.

FIGURE 7: Physical map of cultural facilities in City A.

statistical diagram of the environmental comfort evaluation of cultural facilities in the main urban area of City A.

The five factors influencing the configuration of cultural facilities mentioned above include cultural activities. The development of cultural activities is of great significance to the improvement of residents' utilization rate and the creation of cultural atmosphere. In addition, the development of cultural activities is closely related to the measurement of the service level of cultural facilities, describing from the side whether the hardware and software of cultural facilities can support the development of cultural activities. The evaluation of this indicator is based on the participation of users and the cultural activity data provided by the staff. The research set up five levels in the questionnaire: good, good, fair, not very good, and very bad. Corresponding the evaluation results to the scores of each level in the questionnaire, calculate a quantifiable score.

The activities of cultural facilities in City A mainly include knowledge lecture activities carried out by book-reading facilities, calligraphy and painting exhibitions and publicity activities carried out in cultural and cultural exhibition facilities, festivals, cultural performances, and cultural competitions carried out in cultural activities facilities. Figure 9 is a statistical chart of the evaluation of the development of cultural activities in the main urban area of City A.

Residents' cultural activities are generally organized by cultural departments and spontaneously formed. Activities in large-scale cultural venues are mostly in the form of lectures and exhibitions, and activities organized by

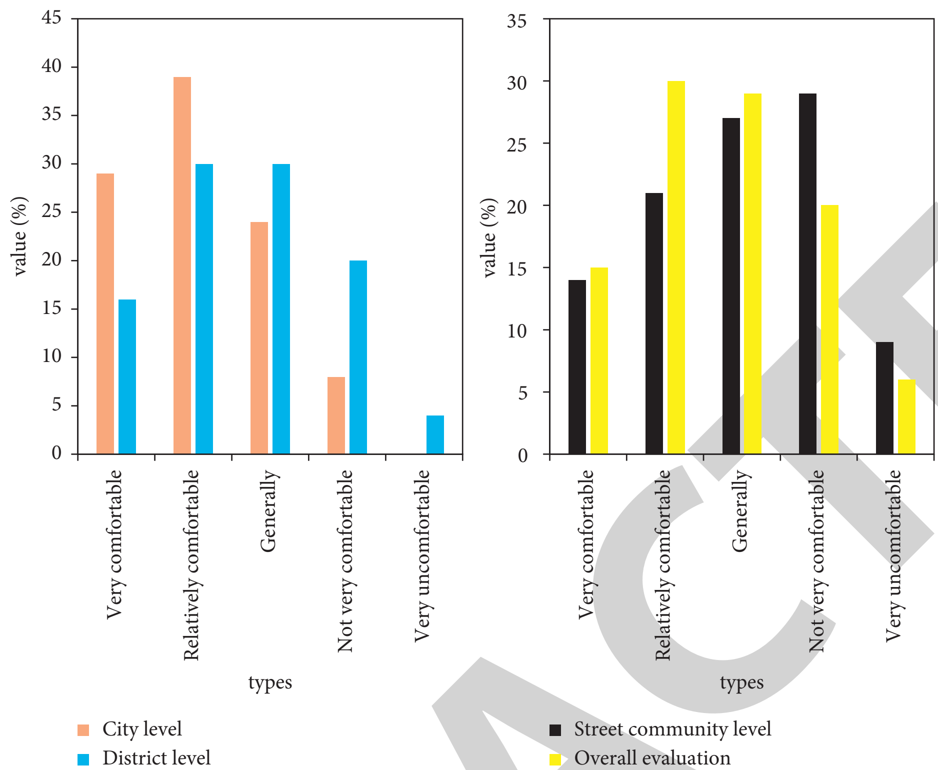


FIGURE 8: Evaluation of the environmental comfort of cultural facilities in the main urban area of City A.

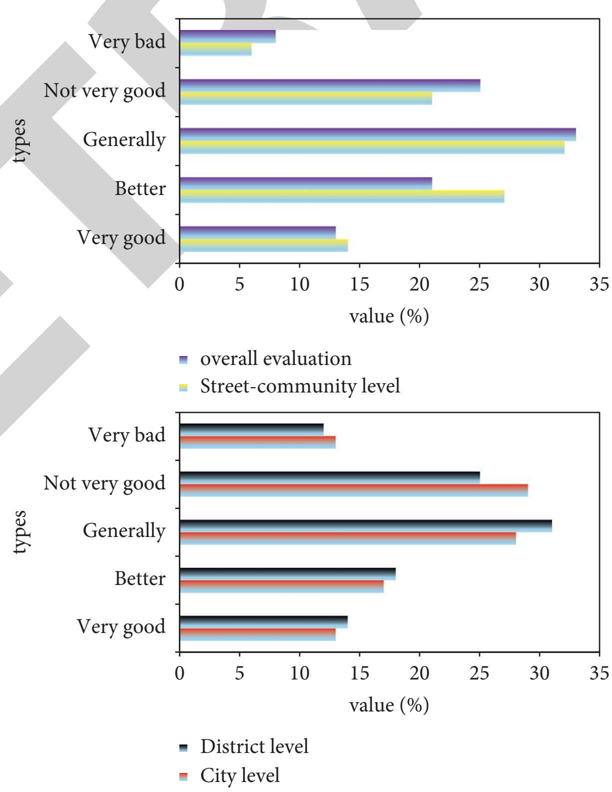


FIGURE 9: Evaluation of the development of cultural activities in the main urban area of City A.

grassroots cultural venues are mostly carried out in the activity venues of streets and communities or open spaces in the community. In the in-depth interviews at various research sites, it was found that residents have a strong demand for cultural activities, but at this stage, many cultural facilities are not equipped to hold large-scale cultural activities. Among them, the municipal and district-level cultural facilities have the equipment and conditions to hold rich cultural activities, but the publicity of cultural activities is not good, and ordinary citizens have a weak sense of participation. Activities such as art competitions are mostly participated in through cultural groups; the form of knowledge lecture activities is single and the types are insufficient. The cultural activities held in the grassroots cultural facilities are closer to the community residents, such as traditional festival celebrations and performing arts activities, but the activities carried out have a greater relationship with the management level of the community, and the situation of different communities is quite different.

Among them, the users of cultural facilities in the main urban area of City A include urban residents and tourists from outside the city. The cultural facilities participated in by tourists from outside the city are mainly cultural and museum facilities supported by historical and natural cultural resources and landmark large-scale cultural exhibition venues. Their main purpose is to visit mainly, is a one-time use behavior, and has greater mobility. Although the number of users of cultural venues has been greatly increased, it is not considered in this paper. This article mainly considers the use of cultural facilities by urban residents.

City-level cultural facilities have rich functions, comfortable environment, and strong attractiveness, so the residents use the composition that is relatively rich and residents of all ages have. At the district level and below, the majority of users of cultural facilities are old men and young people. Young people are more willing to engage in consumer cultural and entertainment activities, and middle-aged people use it less frequently during working days. Within 20 minutes of travel distance, residents are more willing to go to higher-level cultural facilities for cultural activities. Figure 10 shows the evaluation of the utilization rate of cultural facilities in the main urban area of City A, where 1 represents very good, 2 represents good, 3 represents fair, 4 represents not good, and 5 represents very bad.

In actual research, it was found that the actual use of cultural facilities differed significantly. Among them, the utilization rate of city-level cultural facilities scored 77, the utilization rate of district-level cultural facilities scored 72, the utilization rate of street-community cultural facilities scored 69, and the overall evaluation score was 70. Among them, municipal-level cultural facilities are generally more efficient due to their relatively complete functions, and municipal-level cultural facilities can attract residents who are far away. The use efficiency of district-level cultural facilities is closely related to its own hardware configuration, and the use efficiency varies greatly, and the use efficiency gap of street-community-level cultural facilities is relatively obvious. Among them, the use efficiency of street-level cultural facilities is higher than that of community cultural sites, and the use rate of cultural facilities in high-end residential areas is higher than that in residential areas with lower construction levels.

3.2. The Evolution Law of Urban Spatial Morphology Based on Particle Swarm Optimization Algorithm. The spatial expansion and evolution of City A are the result of the mutual influence of many factors. According to relevant data, the topography, location, transportation, economy, policy, and other factors of the city have jointly affected the spatial expansion of the city. Economic development will inevitably lead to urban expansion. At the same time, economic development also needs the support of urban land. According to the regression analysis of the total fixed asset investment, GDP, gross industrial production value, tertiary industry gross production value, and construction land of City A in the urban area of City A, it is found that the total fixed asset investment, GDP, industrial production value, tertiary industry production value of the whole society, and the construction land area of City A all show a certain positive correlation, as shown in Table 1. This means that there is a close connection between economic development and urban spatial expansion.

Selecting the social and economic conditions of City A in recent years for data statistics and using multiple stepwise regression modeling to study the degree of influence of economic and social factors on urban spatial expansion and evolution, the dependent variable is the area of the urban built-up area.

The explanatory variables include 13 indicators: urban population, GDP, GDP per capita, total investment in fixed assets, total retail sales of consumer goods, industrial production, secondary industry production, urban road construction length, road construction area, and real estate investment total, the total sales of commercial housing, the number of students in colleges and universities, and the GDP of the tertiary industry.

It can be seen from Table 2 that since the value of the R^2 coefficient increases with the increase in the number of independent variables in the regression equation, the value of the R^2 coefficient cannot be used as an indicator to reflect the accuracy of the model. However, the modified R^2 value is not directly related to the number of variables, so it can be used as an index to judge the fitness. Combining R^2 and modified R^2, it can be seen that the regression equation has a relatively good fit.

Table 3 shows the results of the analysis of variance at each step in the regression fitting process. It can be seen from the table that when the regression equation contains different independent variables, the significance probability values are all less than 0.001, so the regression equation should include these 4 variables.

Table 4 shows the results of each step of the regression equation process. The significance levels of the constant term, the number of college students, and the total real estate investment are less than 0.001, and the significance levels of the urban population and the total sales of commercial housing are 0.005 and 0.009, respectively. The fitted equation

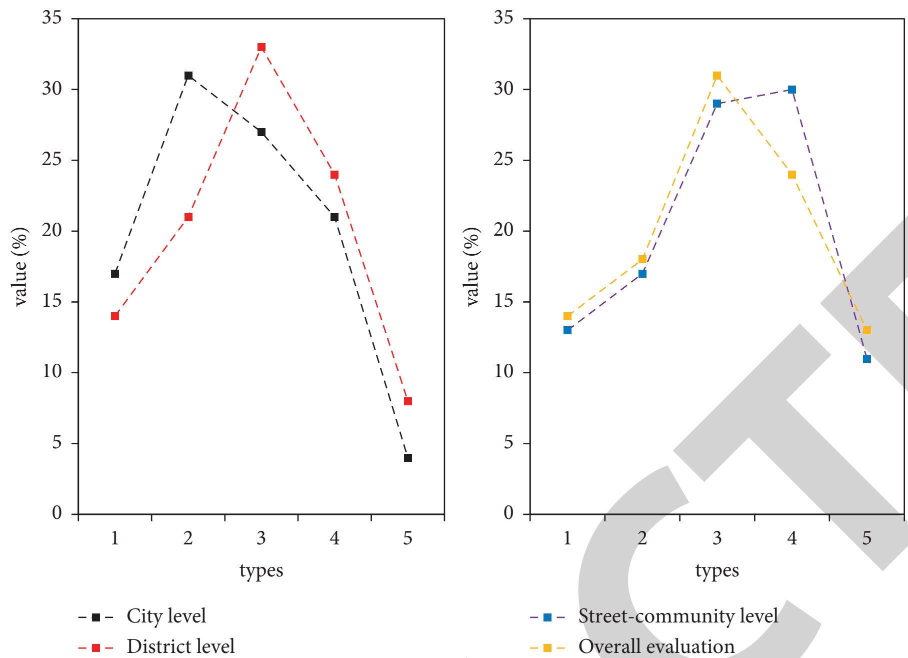


Figure 10: Evaluation of the utilization rate of cultural facilities in the main urban area of City A.

Table 1: The relationship between the economic development and construction area of City A.

Total investment in fixed assets of the whole society (10 million yuan)	Construction land area (km^2)	GDP (ten million yuan)	Construction land area (km^2)	Industrial production value (10 million yuan)	Construction land area (km^2)	GDP of the tertiary industry (10 million yuan)	Construction land area (km^2)
0	100	0	0	0	100	0	100
10000	350	20000	340	20000	400	10000	340
20000	420	40000	530	40000	490	20000	440
30000	500	60000	700	60000	520	30000	520

Table 2: Summary of the fitting process.

Model	R	R-square	Adjust R-squared	Error in standard deviation estimate
1	0.991	0.982	0.981	17.0001
2	0.999	0.999	0.999	4.4574
3	1	0.999	0.999	3.9863
4	1	0.999	0.999	3.4967

Table 3: Analysis of variance.

Model		Sum of squares	df	Mean square	F
1	Regress	360236.213	1	360785.361	1256.734
	Residual error	6602.501	23	291.326	—
2	Regress	367089.736	2	179368.986	8796.501
	Residual error	449.967	22	19.986	—
3	Regress	369782.152	3	119867.961	7308.496
	Residual error	349.137	21	17.062	—
4	Regress	359876.787	4	89978.867	7362.612
	Residual error	251.41	20	11.698	—

Table 4: Regression coefficient table.

Model	Nonstandardized coefficient		Standard coefficient	T
	B	Standard error	Beta	
Constant	89.8653	5.936	—	15.023
Number of students in institutions of higher learning	0.003012	0	0.909	40.002
Total real estate investment	0.050121	0.004	0.291	13.061
Urban population	−0.086312	0.031	−0.079	−3.098
Total sales of commercial housing	−0.013271	0.004	−0.078	−3.012

satisfies the requirements of linearity and homogeneity of variance and has a good fitting effect, which can be used to explain the internal driving mechanism of urban spatial expansion and evolution.

This section analyzes the internal mechanism of urban spatial expansion from the aspects of economy, population, and transportation development and selects 13 influencing factors to analyze the influence mechanism of urban spatial expansion and evolution. Taking the area of urban construction area as the dependent variable and 13 influencing factors as variables, the stepwise regression fitting calculation is carried out. From the equation obtained, it can be seen that the expansion of urban space is mainly linearly related to factors such as the number of students in colleges and universities, the total real estate investment, the urban population, and the total sales of commercial housing. This shows that the number of college students, total real estate investment, urban population, and total commercial housing sales have a significant impact on the expansion and evolution of urban space.

4. Discussion

There is a close relationship between urban cultural strategy and space construction, because urban cultural strategy often must have a load on the medium to be effectively expressed, and one of the main manifestations of space load is the urban cultural landscape space. In fact, the spatial change of the urban cultural landscape is also the most intuitive effect in the urban cultural strategy. According to the scale classification of urban space, the spatial effect is mainly manifested in the emergence of historical and cultural gathering areas, the renewal of historical and cultural spaces, and the increase of historical and cultural spaces in isolation and historical and cultural landmark systems. Urban spatial structure refers to the interrelationship of important components such as the material environment, social function activities, and human values in the city. The expansion and evolution of City A's morphological structure are affected by natural conditions, traffic factors, economic development, population growth, government regulation, and other factors. This paper selects 13 variable indicators such as urban population, GDP, and GDP per capita from the aspects of transportation, population, and economy and uses the stepwise regression analysis method for in-depth study and analysis of the impact mechanism of 13 factors on urban expansion and evolution. From the results of multiple regression analysis, the number of students in urban colleges and universities, the total real estate investment, the urban population, and the total sales of commercial housing have a significant impact on the expansion and evolution of the urban spatial structure. Among them, the number of students in urban institutions of higher learning is the most important factor driving the spatial expansion and evolution, followed by the total real estate investment, urban population, and total commercial housing sales.

5. Conclusions

In the context of the global economic and cultural globalization of the world, cultural soft power has become a concentrated expression of the overall strength of the economy and cities.

At the same time, in order to respond to the development needs of the cultural industry and establish a public service-oriented government, public and benefit-oriented urban cultural infrastructure planning will be adopted to fill the gaps in the contradiction between supply and demand in the construction of cultural industries. At the same time, in order to achieve the fairness and efficiency of the urban cultural infrastructure planning and construction process, research and analysis have been carried out on the evaluation of the urban cultural infrastructure configuration performance. It is found that the actual use of cultural facilities is significantly different, and the municipal cultural facilities are generally more efficient due to their relatively complete functions. In addition, municipal cultural facilities can attract the use of residents who are far away. The use efficiency of district-level cultural facilities is more related to its own hardware configuration, and the use efficiency varies greatly. In this regard, it is particularly important to eliminate the difference in the use of cultural facilities between urban and rural areas. Carrying out the evaluation of the configuration performance of urban cultural infrastructure is of great significance for understanding the current situation and determining the direction of urban planning. Based on the particle swarm algorithm, this paper simulates the cultural space users as a flock of birds and the urban cultural space as food and simulates the behavior of seeking cultural facilities as birds searching for food. Based on the results of this survey and research, the utilization rate of public cultural facilities in City A is 77, the utilization rate of district-level cultural facilities is 72, the utilization rate of street-community cultural facilities is 69, and the overall evaluation score is 70. The actual use of cultural facilities varies significantly, so it is particularly important to eliminate the differences in the use of cultural facilities between urban and rural areas. Among them, this article is limited by graphic

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9847245, 1 page
https://doi.org/10.1155/2023/9847245

Retraction

Retracted: Machine Learning: The Backbone of Intelligent Trade Credit-Based Systems

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This is article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This is investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] F. Shah, Y. Liu, A. Anwar et al., "Machine Learning: The Backbone of Intelligent Trade Credit-Based Systems," *Security and Communication Networks*, vol. 2022, Article ID 7149902, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7149902, 10 pages
https://doi.org/10.1155/2022/7149902

Research Article

Machine Learning: The Backbone of Intelligent Trade Credit-Based Systems

Faiza Shah,[1] Yumin Liu,[1] Aamir Anwar ,[2] Yasir Shah,[1] Roobaea Alroobaea ,[3] Saddam Hussain ,[4] and Syed Sajid Ullah [5,6]

[1]*School of Business, Zhengzhou University, Zhengzhou 450001, China*
[2]*School of Computing & Engineering, University of West London, London, UK*
[3]*Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia*
[4]*School of Digital Science, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong, BE1410, Brunei Darussalam*
[5]*Department of Information and Communication Technology, University of Agder (UiA), Grimstad, Norway*
[6]*Department of Electrical and Computer Engineering, Villanova University, Villanova, PA 19085, USA*

Correspondence should be addressed to Saddam Hussain; saddamicup1993@gmail.com and Syed Sajid Ullah; syed.s.ullah@uia.no

Received 7 January 2022; Accepted 12 February 2022; Published 8 March 2022

Academic Editor: Chin-Ling Chen

Technology has turned into a significant differentiator in the money and traditional recordkeeping systems for the financial industry. To depict two customers as potential investors, it is mandatory to give the complex innovation that they anticipate and urge to purchase. In any case, it is difficult to keep on top of and be a specialist in each of the new advancements that are accessible. By reappropriating IT administrations, monetary administrations firms can acquire prompt admittance to the most recent ability and direction. Financial systems, along with machine learning (ML) algorithms, are vital for critical concerns like secure financial transactions and automated trading. These are the key to the provision of financial decisions for investors and stakeholders for the firms which are working with the trade credit (TC) approach, in Small and Medium Industries (SMEs). Huge and very sensitive data is processed in a limited time. The trade credit is a reason for more financial gains. The impact of TC with predictive machine learning algorithms is the reason why intelligent and safe revenue generation is the main target of the proposed study. That is, the combination of financial data and technology (FinTech) domains is a potential reason for sales growth and ultimately more profit.

1. Introduction

FinTech is an abbreviation or single word for monetary advances and indicates the collective form of digitized money and information technology innovations. It is utilized especially by banks and financial administrations associations to direct their business tasks more effectively and give better monetary administrations to their clients. It might take the state of programming or an app that empowers organizations to give mechanically complex and contactless administrations to their clients by setting up monetary exchange processes. FinTech, by facilitating convoluted monetary decision making, has significantly changed the banking and monetary administrations industry [1]. Monetary technologies have changed banking and monetary administrations activities around the world throughout the last ten years. They have worked on the clients' and banking specialists' lives fundamentally. Particularly, when we talk about innovation, we are alluding to online exchanges, web banking, banking applications, and online stock exchanging, in addition to other things [2]. Regardless of how much information there is, we could not realistically exaggerate the effect it has on the present economy. There is enough information at our disposal, anything from our mobile phones to online media use, web browsing, and e-commerce exchanges [3]. Although large information and information science have been accustomed to bringing about change in the banking, enormous information applications, and monetary administrations areas, these organizations have had especially sure outcomes in carrying out these changes.

FinTech is being utilized for an assortment of significant monetary capacities like advanced installments, contribution, and abundance the executives, just as loaning and advance reimbursement, exchanging, and individual banking. Personalization, reconciliation, verification, and information following and examination are the useful components of FinTech applications [4].

Financial technology has gained popularity as a relatively new concept since 2010 with the title of FinTech. It is the term used to depict the combination of PC programs and other innovation drives to help banking and monetary administrations. It is a consoling field brought into the world of the combination of advanced stages and man-made consciousness administrations, to manage monetary exercises. AI is having a significant impact on how consultants behave as compared to traditional financial regulatory bodies [5]. In the last decade, a large amount of FinTech trading has had a significant boost in the financial economy. Many venture capitalists are interested in these FinTech businesses, and many scholars have looked at the technological and economic variables that drive venture capital investment. They found that nations with more FinTech firms have more evolved economies and adventure financing accessible. FinTech ventures developed rapidly, coming to a peak from 2008 to 2014, with over twofold capital increase [6]. FinTech is one of the quickest developing areas of the economy, as per an overview by Accenture. FinTech applications and administrations cover a wide scope of points. Online installments and cash moves, loaning, resource, venture executives, computerized banking, and individual accounting are among the themes impacted by FinTech. Different parts of the FinTech business are acquiring consideration step by step. It incorporates InsurTech, advanced venture executives, computerized loaning, installments, advanced banking, and blockchain innovation [7]. The financial sector relies heavily on financial data, which is utmost sensitive data, with enormous amounts of historical and market data emerging.

In the new era where technology is becoming a mandatory part of technological growth, financial systems are relocating towards more versatile and user-friendly environments. The consumer facilitation and ease of understanding are highly considerable reasons which involve click transactional changes as compared to the outdated systems where the person has to go and be physically available for the heave money to be transferred from one location to the other. The financial systems involve accurate and timely transference [8]. Financial transactions are never considered with insecure systems to breach one's identity. The technology ensures every system the accurate and complete transference of information from one part of the region to the other. IoT, smart devices, and e-commerce are some of the examples that provide fast, effective, and user-friendly environments, as well as ensuring secure financial transactions.

Innovation in the monetary area is changing how administrations are conveyed; it is likewise making contenders outside conventional businesses. Computational benefits of late times have permitted the assortment of crucial firm information, similar to continuous exchanges and client information, is of interest to back experts who may have the option to apply it in stock estimating investigation [9]. FinTech is said to lastingly affect the entire business, due to its exceptionally inventive and possibly problematic repercussions. A few definitions consider it an industry, while others characterize it as an innovation. Of course, hardly any different writings characterize it as a kind of activity, like a business or organization. Every one of the sources recognizes FinTech as something novel, emerging, upsetting, and creative. To lay it out simply, FinTech is a clever financial industry that uses advancement to make money-related activities more useful [10]. The speedy extension in interest in features is a requirement for better perception of it. FinTech advancements are completely important to pioneers and the money business. In any case, a couple of FinTech headways can unfriendlily affect explicit money-related endeavors. At the point when such headways come from decently new, nonfinancial firms, they can have more unfavorable results [11]. The fundamental goal of the money-related region is to enable trade. For model, cash, and portion game plans, the spread of splendid device models, etc. The financial sector's and various undertakings' capabilities are frequently obscured by these tendencies. There is a strong fuse between FinTech plans with the fundamental likewise like the discretionary regions. New strategies prompt the FinTech improvement at the association level [12]. FinTech renames the way through which clients save, store, get, spend, set aside, and guarantee cash. There are a couple of FinTech strategies, like overflowing the board models, crediting models, portion models, insurance organization models, and capital market models, that are executed by the extending number of FinTech new organizations. Such startups regularly attempt to rival the current customary monetary foundations, or they attempt to team up with them. When the different financial guidelines are great, they will quite often be more serious and less cooperative.

The main objective of the paper is to prove that the machine learning algorithms are reliable in terms of financial decision making. Whether the SMEs data is treated as supervised or unsupervised, the trade credit plays an enhanced role in the successful regulation of the organization. This ML support system is multidimensional and applicable for financial decision making systems including various terms like data science, AI, and smart business decisions. Finally, the results of ML algorithms indicate the best one that can be selected for business intelligence.

Section 1 was a thorough introduction about the FinTech and trade credit. Further sections of the paper include the following: Section 2 is the literature review. Section 3 is about FinTech and its impact on data with data science. Section 4 is about the machine learning algorithm and its implementation. Section 5 states the experiment setup. Section 6 provides the results and discussion.

2. Literature Review

The financial sector aims for profitable revenue generation systems. The sophisticated systems are using Artificial

Intelligence (AI) for decision making. AI is the technique used to make machines more intelligent. The maximum intelligence level of human beings is considered ideal since they can solve many problems with a finite number of achievable solutions. The most important concern is achievability and accuracy [13]. For this purpose, in FinTech, the systems are programmed to generate accurate results. These algorithms are goal-oriented, instead of being data-oriented. The machine learning AI-based algorithms are executed for supervised techniques [14] where the complete and labeled data set is provided. The other concept is unsupervised techniques where the data set is unlabeled and the machine has to think and decide.

Algorithmic exchanging is a term that alludes to the act of utilizing calculations to settle on more educated exchanging decisions. Ordinarily, merchants utilize numerical models that constantly screen organization news and exchange action for any factors that might make share costs increase or decline. The model [15] comes pre-customized with guidelines on various viewpoints—like time, value, amount, and different factors—for naturally making exchanges without the broker's dynamic support. In contrast with human dealers, algorithmic exchanging can assess tremendous measures of information simultaneously and consequently execute a huge number of exchanges every day. AI empowers merchants to settle on fast decisions, giving them an edge over the market normal.

Information mining [16] is the center of information revelation in the data set, which analyzes huge existing data sets to foster the models which are utilized to do examination and expectation. Hence, information mining is utilized to observe the examples of explicit informational collections and to take on the methods created by ML for expectations.

ML trains models by taking care of gigantic, quality existing information, gaining from the information to settle on a learned choice without programming [17]. Along these lines, AI involves ML for its wise conduct. ML is a gigantic piece of AI. For ML and data mining, ML revolves around the improvement of computations through iteratively dealing with planning educational assortments and the estimates of the outcome. Curiously, data mining is based on finding models and examples in the current data. Thus, data mining systems fill in as commitments for ML, while ML takes on data mining estimations to set up models [18].

FinTech is gradually empowered in various financial-technical sectors. throws light on financial setups under the study of Artificial Neural Networks (ANNs) [19]. Decision learning support systems with ANN exactly use the knowledge discovery deduction, as a human brain computes the input to convert it into respective output. ANN involves complex problem-solving techniques by applying multiple layers, thus matching the preliminary results from the first layer till the final one; some of the layers are also called hidden layers. Financial systems like image recognition, voice recognition, and biometric verification systems are all concluded under the umbrella of ANN techno-financial systems [20]. The utility of ANN is feasible for such systems where the determination is the conversion from a linear to dynamic problem-solving technique for the financial sector.

Information researchers, as a rule, embrace administered learning methods or unaided learning procedures to recognize charge card deceitful exchanges [21]. Regardless, a couple of researchers have proposed the unsupervised learning procedure to perceive charge card counterfeit trades, Figure 1, as FinTech data-driven strategy structure [22]. The instructive assortment from the machine learning method is uncommonly imbalanced, with the extent of 0.00173 between the amount of deception and authentic trades. Therefore, the Cluster Centroid technique [23] is used to under-model this imbalanced enlightening assortment.

The studies [24,25] further concentrated on the execution of the KNN calculation and anomaly discovery strategies for the improvement of answers for recognizing Visa misrepresentation. Their review demonstrated that KNN is quick with the least bogus cautions. The KNN strategy is exact and productive, as affirmed by the analysis results. In the first place, the fluffy rationale is designed to communicate mental vulnerabilities. It estimates the strength of significant worth for probabilistic measures or unquantifiable measures somewhere in the range of 0 and 1.

A computation considering fuzzy ID3 was proposed [26] to perceive charge card distortion. In this paper, the maker presents the technique associated with building an ID3 decision tree using cushioned reasoning and applies the fleecy ID3 tree on an educational record. For charge card coercion revelation, the overall precision is used to check the show. Nevertheless, deception getting rate (True Positive Rate) and fake alert rate (False Positive Rate) are better estimations while surveying the learned blackmail classifiers [27]. The above assessment showed that RF performs best with high precision and coercion getting rate. Likewise, we saw that the data assortment is by and large astoundingly imbalanced; oversampling and undersampling systems are customarily embraced in tests for better execution.

3. FinTech for Sensitive Data

FinTech, despite its reputation as a young sector, has a long history that may be broken down into three periods. FinTech includes the creation of mainframe computers, SWIFTS, ATMs, and other financial technology [28]. FinTech was the name given to the next phase of financial technology, which included the Internet and the Internet of Things [29]. We are currently in a transition period of FinTech where more and more technologies are projected to emerge. Finance encompasses all aspects of financial management, as well as technology management and innovation management.

To make financial operations more efficient, the deployment of technological solutions in novel ways is the common observation for the upcoming era. As a result, FinTech is a multidisciplinary topic [30]. A. FinTech Steamships telegraphs, and railroads all allowed for better financial links between countries. FinTech is frequently thought to be new. The association would slice the time it takes to impart between North America and Europe from as long as 10 days to 17 hours. This improvement laid the foundation for the rise of a few incredibly fruitful protection, banking, and joint-stock undertakings, exceptionally important to modern upgradation.

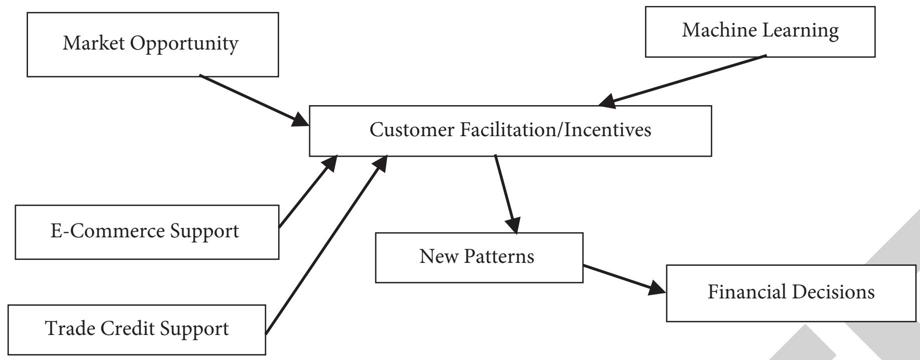


Figure 1: Machine learning for financial decisions.

The revised scheme acts as a foundation for financial globalization. Several technological advancements occurred in the last decades. The risks of computer management systems were first disregarded, as seen by the failure of Long-Term Capital Management following the financial crisis of 1997–98 [31], and the Internet's arrival paved the way for FinTech [32]. E-banking introduced additional hazards, particularly for regulators, since technology enabled fast money withdrawals. Authorized financial organizations, such as banks, were intended to be the suppliers of these e-banking services. With every passing day, more people trust technology companies to manage their money more than traditional banks. Exceeding millions, loaning stages have been laid out in evolved nations, which work outside of any managed structure globally [33]. The way that they work outside of the administrative structure does not seem to upset moneylenders or borrowers, who are more worried about low expenses, more straightforwardness, and greater returns.

3.1. Global ICT for Financial Refinements. Banks' competitiveness and profitability were seriously harmed, and laws and procedures relating to stress testing, among other things, only added to the already increasing costs. In the case of Application Programming Interfaces (APIs), FinTech required a high level of smartphone penetration and sophistication. Another significant phase of FinTech is characterized by the rapid development of technology and a shift in the identity of financial service providers [34,35]. Financial institutions are at risk of having their services delivered to consumers and businesses in a more targeted, improved, and convenient manner by startups and IT firms. Figure 1 indicates the involvement of machine learning algorithms in facilitating financial decisions based on machine support systems. The input data is processed under critical rules, and thus the results always determine the profit or loss ratio [36,37]. The new hidden patterns from data that were previously unknown are now easily determined.

The business domains are continuously improving due to IT and various applications developed due to software. This aspect can be seen in Figure 1. The central concern is customer facilitation. The financial activities should be seamlessly activated. The customer invests by observing market opportunities. Advanced trading like e-commerce that regulates online systems is the user-friendly system that undergoes continued progress of the financial regulations. Trade credit is an open hand for the customer. It means the investor is an entity, and the buyer is another independent entity, but the facilitator is the intermediate entity who offers services with freedom of financial load. Thus, we say time is money.

FinTech focused its restrictions on systemically important institutions. However, it is now considered vital to begin focusing on certain industry participants. FinTech's evolution has resulted in parallel advances in regional technology enhancement. It is claimed that a multilevel and flexible strategy is appropriate and that regulations should be enforced in varying degrees depending on the size and risk of companies [38]. Regulators would have to collaborate with industry to obtain clarity of the changing market and establish policies that stimulate innovation while also stabilizing risk and eliminating the possibility of regulatory arbitrage [39]. In the financial sector, technology is both revolutionizing how services are offered and creating competition outside of established industries. Finance experts who might have the option to use it in stock valuing examination will be interested in computational advances that have lately permitted the gathering of major firm information, for example, continuous exchanges and client information [33]. Because of its profoundly imaginative and possibly troublesome implications, FinTech is said to affect the whole area [40]. Others arrange it as an industry by certain definitions and as an innovation. Hardly any different sorts of writing, then again, characterize it as a type of activity, like a business or help.

3.2. Credit Decisions with Data Mining. FinTech is associated with knowledge discovery in databases (KDD), which is technically data mining [41]. It examines massive existing databases to construct models for analysis and prediction. As a result, information mining is utilized to find designs in explicit informational indexes and to apply ML strategies to make forecasts. ML trains models by giving them a lot of excellent existing information and gaining from it to settle on informed choices without the requirement for programming. Therefore, AI's intelligent behaviors are based on machine learning [42]. AI includes a lot of machine learning. In terms of machine learning and data mining, machine learning focuses on improving algorithms by iteratively contributing preparation informational indexes and making expectations about the result.

Information mining, then again, centers around recognizing examples and patterns in existing information. Thus, information mining strategies are utilized as a contribution to ML, and ML utilizes information mining calculations to make models [41]. Fake neural organizations are utilized to impersonate the human mind in profound understanding, which is utilized in information mining to deal with complex issues like picture and voice acknowledgment. In the monetary region, variable advances work on the woods and examination process. With organized and dynamic model properties, DL shifts observational to nonlinear examination. Man-made consciousness is a famous term that alludes to robots' capacity to imitate human conduct and make decisions similarly to people [43]. Accordingly, rather than depending on learning, AI frameworks are expressly composed. Information mining strategies are the foundation of AI, and information mining fills in as programming codes for the information and data expected by AI-controlled frameworks [44]. Therefore, information-driven methodologies allude to any innovations that are information-driven and depend on the information-driven independent direction or system creation.

4. Machine Learning Algorithms Support

Artificial Intelligence is regulating intelligent decision making by using machine learning algorithms. Recently, a more noteworthy number of scientists have shown the amazing exact exhibition of ML calculations for resource value estimation when contrasted and models created in customary measurements and finance [45,46]. The capacity of an individual or firm to all the more precisely gauge the expected cost of any resource has colossal worth to professionals in the fields of corporate money, technique, private equity in expansion to those in the fields of exchanging and speculations [47]. Lately, banks additionally endeavored to use ML techniques for strategy investigation and macroeconomic direction.

ML calculations and methods should remove designs gained from historical data in interaction and thus make exact expectations on new information [48]. The appraisal of the accuracy of our calculations is the testing. While there exist a huge number of types and classes of ML calculations [49], a high level of the exploration papers in the current academic literature outlines the issue of monetary resource value gauging as administered learning issue. Given the functional and exact focal point of this paper, calculation definitions and numerical verifications of algorithms will be overlooked.

4.1. Data Science FinTech Businesses Horizons. With new-age information science and man-made consciousness instruments, finance has become more intuitive. FinTech [50] is in the focal point of ML approaches [51] integrating, enhancing, and changing monetary administrations, economy, innovation, media, correspondence, and society. Rationale, arrangement, information portrayal, display, independent frameworks, multiagent frameworks, intricacy science, master frameworks, choice emotionally supportive networks, advancement, reproduction, design acknowledgment, picture handling, and regular language handling are for the most part instances of AI for monetary frameworks spine.

FinTech-based organizations manage astute distinguishing proof and validation, a security protecting handling [52], progressed portrayal learning, progressed investigation and learning, information disclosure, computational knowledge, occasion, and conduct examination, and web-based media and organization examination; and later advances, for example, profound learning [53], computerized cooperations, learning, and reactions, and complex statistical and numerical displaying are two additional fundamental disciplines.

FinTech has the characteristics of cross-market and cross-industry and leads to the diversification of the financial service market. However, it does not change the traditional purposes and security principles of any financial activity and brings new challenges to the regulation of the traditional financial industry [54]. It is hard for the customary administrative systems to manage the likely dangers of monetary advancements driven by arising innovations, like AI, blockchain, distributed computing, and large information. These variables give FinTech organizations more motivating forces for resistance advancement. The real factors of inordinate advancements and inadequate guidelines lead to noticeable FinTech hazards. For example, cash loans and campus loans caused excessive borrowing, violent collection, extremely high rates, infringement of personal privacy, and many other issues [55]. Speculation in virtual currency led to price spikes and severely disrupted financial order.

Business regions in FinTech incorporate areas, which depend on ML approaches, covering every aspect of a monetary framework and its environmental factors, just as all monetary firms. The significant business regions in savvy FinTech are monetary and financial developments, for example, new instruments and monetary business sectors (which incorporate items and administrations), financial monetary members (which incorporate retail and individual financial backers, organizations, and controllers), and financial monetary practices.

Hazard overseeing and upgrading computerized monetary standards including digital forms of money; making, getting, assessing, hazard making due, and streamlining portable financial organizations and administrations; customizing, robotizing, approving, getting, hazard making due, and enhancing Internet/web-based banking; empowering, customizing, mechanizing, getting, hazard overseeing and advancing open banking; assessing, getting and hazard overseeing shadow banking [56]; and creating more astute and more mechanized, customized, versatile and drawing in financial administrations; and so forth savvy protection empowers protection items, frameworks, and administrations to guarantee safe/secure, practical, proactive, custom-made, trustful, strong, secure, and hazard opposed wellbeing, vehicle, home/content/building, travel, and different organizations.

Individual and business protection items and administrations are among the ML research themes, in addition to empowering early, dynamic, and advancing protection extortion discovery; making dynamic, customized, and time-shifting protection item and administration suggestions; assessing, breaking down, recognizing, making do with, and enhancing protection hazard and consistency; assessing, robotizing, distinguishing, and improving protection security; and making novel protection items and administrations.

For people, partnerships, or activities, shrewd loaning empowers hazard unwilling loaning, credit, and home loan items and administrations that are customized, focused, prescient, effective, versatile, and secure [57]. The ML research headings for shrewd driving cover ventures and assignments—for example, settling important angles and issues through blockchain; robotizing crowd funding, like mission creation and system advancement; and tending to significant viewpoints and issues through man-made reasoning.

Each part of FinTech requires continuous development and exploration, which is the place where Al innovation proves to be useful. Information and learning-empowered frameworks and administrations for mechanized evaluation, credit scoring, advance valuation, exchanging technique age, client chatbots, monetary preparation, security alarming, consistence moderation [58], proactive and customized crowd funding project suggestions, cross-item advance, protection, venture portfolios and estimation, and custom-made danger alleviated frameworks and administrations for resource and abundance portfolios are instances of AI-driven FinTech developments.

Enhancement and hazard the board for high-esteem clients; distinguishing and mediating in untrustworthy and uncertain exchanging, loaning, credit and advance valuation, installment, promoting, contest, and guideline; and giving entire of-business [59], security saving, and united FinTech and EcoFin organizations and administrations to huge scope, circulated and associated networks or social orders, in addition to other things. All of the above brilliant FinTech organizations and regions, then again, share a few fundamental cycles, capacities, and exercises.

4.2. Machine Learning for Smart Business Decisions. The AI-based networks analyze the direction of the period of shrewd FinTech and likely open doors for brilliant FinTech fates. The shrewd FinTech time has shown up; savvy FinTech is acquiring a foothold and turning into a basic part of the present and the upcoming economies, social orders, and advancements [60]. Through channels, for example, QR codes, WiFi organizations, portable applications, online media organizations, short informing stages, smart advanced partners, and the Internet, FinTech interfaces with each individual, association, item, administration, and action anyplace, anytime, and in any structure. They hoard an enormous number of substantial and elusive resources and administrations, just as an assorted scope of items, applications, and administrations.

Simulated intelligence and information science are turning out to be more significant in making finance more astute and advancing FinTech's consistently developing cleverness, just as the insight of independent monetary frameworks and custom-made monetary administrations. Canny ID and confirmation, independent associations and interchanges, cloud examination, profound learning [61], unified learning, cross-market investigation, profound monetary demonstrating, and robotized collaborations and reactions are for the most part instances of shrewd innovation progressions that are advancing to address the arising difficulties and amazing open doors in more intelligent monetary organizations and greater monetary information. Security is one more tremendous worry for the monetary administration industry. Information breaks include monetary administrations firms [62]. Cybercrime presently costs the monetary administrations are more than some other industry seemingly forever.

The fast ascent of information science as an expert field has trickled in individuals from all foundations. Engineers, PC researchers, promoting and finance graduates, examiners, human asset faculty, and everybody need a piece of information science; that consistently experiences the capable business knowledge (BI) experts hoping to land their first information science job. They are frequently disappointed by the apparent absence of chances for them. A ton of them feel that their job is dull, or they simply need to carry out whatever is asked of them. They miss being nearer to information science openings than some other experts out there. BI experts hold a gigantic benefit over nearly anybody attempting to progress in information science given the accompanying reasons: BI experts as of now approach information researchers in different ventures [63]. Often we use traditional database management systems to know about hidden patterns. This mechanism is useful for small or tangible data but sophisticated systems are moving towards big data; for this, we need to apply smart or AI-based techniques. Thus, the results are not only clear but also deterministic. BI experts have the business setting, and they work intimately with organizations. They have insight with fundamental information investigation ventures as frequently business requests these notwithstanding the reports they use.

FinTech's rapidly rise to prominence as a global leader. It is found that additional advantages from resistance development, prizes from consistent advancement, and administrative power punishment force impact FinTech organizations' essential choices, while administrative expenses, social assessment, and negative externalities impact administrative power key choices. At last, strategy proposals are exclusively founded on current realities of unreasonable advancement and inadequate guideline as FinTech business.

5. Software Environment

RapidMiner is chosen to execute the machine learning algorithms. The experimental setup is executed on core i-7, using a desktop system, including 4 GB RAM and 64-bit Windows 10 Home edition operating system.

RapidMiner is an efficient statistical tool. It is used for data preprocessing, anomaly reduction, and data quality enhancement. The data cleansing is applied in a user-friendly and interactive layout. The refined data is ready to use for the actual processing.

5.1. Data and Methodology. The data selected for the proposed study is nonfinancial firms' data from Pakistan over the years 2017 to 2021. The data consists of 1357 companies that are functional in favor of trade credit-based nonfinancial systems. The data is preprocessed and refined with complete and accurate values. The clean data is further processed for machine learning algorithms, applying the systematic approach of Figure 2. It is technically important that the machine learning techniques were used in both supervised and unsupervised machine learning algorithms. Both approaches are chosen to indicate that the predictive patterns are determined. The supervised learning with labeled data along with the trade credit approach is beneficial for results generation.

FinTech is expanding into a big family that includes banking tech, trade tech, lending tech, insurance tech, wealth tech, payment tech, and risk tech. As depicted in Figure 2 which portrays a multifaceted scene of brilliant FinTech and the amalgamation of these parts, the savvy FinTech environment is multilayered cooperative energy between IT-based business targets, business regions, information, assets, and supporting innovation. By and large, brilliant FinTech environments can be partitioned into far-reaching FinTech-driven ventures, areas, cycles, capacities, and exercises.

The essential cycles and their critical capacities and exercises are portrayed in Figure 2: plan, produce, work, advance, streamline, and shield. From one viewpoint, FinTech-driven undertakings and enterprises are tremendous and quickly evolving. The FinTech system is portrayed as above. Every one of these key areas is additionally portrayed by the organizations and the strategies that help it.

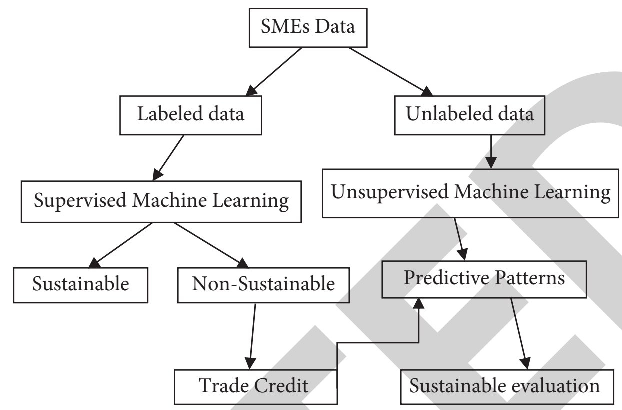


Figure 2: The machine learning-based predictive model.

Table 1: Financial tradeoff analysis.

	AUC	F1 score	Precision	Recall
Neural Network	0.86	0.812	0.8421	0.812
CN2 rule induction	0.81	0.74	0.7612	0.74
Random Forest	0.79	0.737	0.7451	0.737
Logistic Regression	0.72	0.64	0.6531	0.64
Naive Bayes	0.6	0.57	0.5931	0.57
SVM	0.69	0.65	0.6637	0.65
AdaBoost	0.72	0.67	0.6912	0.67
KNN	0.45	0.39	0.4187	0.39

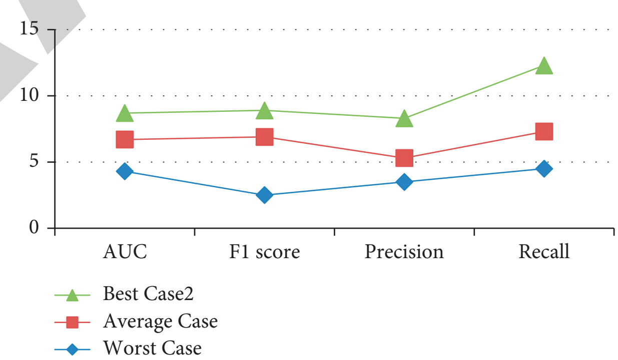


Figure 3: Results of ML algorithms.

6. Results and Discussion

It is vital that frequent preprocessing and information purifying erase absent and strange or mistaken qualities. The data items with the highest value are subjected to preprocessing. The missing characteristics were replaced with the most possible/nearly evaluating expected characteristics. The useless data were seen to be the least effective, that may be ignored, which is exceptionally less in degree. The issue of overfitting and underfitting was carefully seen, so the data quality should be consistent. The data in the wake of preprocessing is quality-arranged. Our goal is to encourage effective FinTech regulation and strike a balance between regulation and innovation, which are presented in Table 1. The ML-based algorithms like Neural Network, CN2 rule induction, Random Forest, SVM, AdaBoost, KNN, Logistic Regression, and Naive Bayes were used to determine the results. Table 1 shows promising results that are represented by the AUC, precision, and recall and in terms of accuracy of the system.

The danger control components of some P2P loaning stages were not sound. To advance the great improvement of FinTech, it is critical to fortify the viable guidelines of FinTech developments and accomplish harmony among guidelines and advancement. In the development and innovation of the FinTech industry, whether FinTech companies choose compliance innovation is a complex game process, and the formation of compliance willingness depends on certain policy incentives and guidance. Figure 3 shows the results with ML algorithms classified into three categories. The precision, accuracy, and F1 scores are not manually possibly determined. The training and testing models are the specialties of AI support systems. Considering these decisions are refined and more reliable as compared to traditional systems.

The choice of objectives, the heterogeneity of cognitive ability, and the complexity of the economic environment between different players not only determine their characteristics of bounded rationality in the game process but also make the game process full of dynamics. The theory of the evolutionary game is just an important means to study the dynamic game relations between different players. Subsequently, it is sensible to build a developmental game model to examine the game conduct of the primary members in the FinTech market.

7. Conclusion

Information systems guarantee financial data security and increase its utility. The proposed study is in favor of IT infrastructure for intelligent financial decision making systems. The SMEs are less sustainable industries but are slow progressive industries. On the basis of the late payment technique, trade credit favour can help SMEs to survive. The fewer financial sources, exchange, and sale of goods are the sustainable parameters of micro-financing preservation. ML algorithms can determine predictions about the decision support for regulating financial decisions. The system's reliability for historical five-year data is applied through ML algorithms, and the results show that the best case organizations have more chances of survival and may lead to a progressive approach [64].

Data Availability

The Data used in this research can be obtained from the corrosponding authors upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are grateful to the Taif University Researchers Supporting Project no. TURSP-2020/36, Taif University, Taif, Saudi Arabia.

References

[1] E. Basuno, Y. Yusdja, and N. Ilham, "Socio-economic impacts of avian influenza outbreaks on small-scale producers in Indonesia," *Tran boundary and Emerging Diseases*, vol. 5, no. 1-2, pp. 7–10, 2010.

[2] F. Shah, Y. Liu, Y. Liu, Y. Shah, and F. Shah, "Influence of shareholder equity on trade credit demand: The study of non-financial firms in Pakistan," *R-Economy*, vol. 7, no. 1, pp. 61–67, 2021.

[3] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," *Computational Economics*, vol. 57, no. 1, pp. 203–216, 2021.

[4] F. Shah, J. Li, Y. Shah, and F. Shah, "Broad big data domain via medical big data," in *Proceedings of the 2017 4th International Conference on Systems and Informatics (ICSAI)*, IEEE, Hangzhou, China, 11–13 Nov. 2017.

[5] G. Babaei and S. Bamdad, "A new hybrid Instance-based learning model for decision-making in the P2P lending market," *Computational Economics*, vol. 57, no. 1, pp. 419–432, 2021.

[6] F. Shah, J. Li, W. Zhou, J. Khan, and F. Shah, "Hybrid Compression of medical images with Wavelets," in *Proceedings of the 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, IEEE, Chengdu, China, 14–16 Dec. 2018.

[7] Y. Shah, Y. Liu, Y. Liu, F. Shah, and F. Shah, "Challenges of small- and medium-sized businesses in Pakistan due to COVID-19 pandemic," *R-Economy*, vol. 6, no. 3, pp. 222–226, 2020.

[8] F. Shah, J. Li, Y. Shah, and F. Shah, "Retrospect and prospect of wavelet and big data: a survey," in *proceedings of the 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, IEEE, December 2016.

[9] N. Paltrinieri, L. Comfort, and G. Reniers, "Learning about risk: machine learning for risk assessment," *Safety Science*, vol. 118, pp. 475–486, 2019.

[10] J. Sen and S. Mehtab, "A comparative study of optimum risk portfolio and eigen portfolio on the Indian stock market," *International Journal of Business Forecasting and Marketing Intelligence*, vol. 7, no. 2, p. 143, 2021.

[11] Y. Lei, Q. Peng, and Y. Shen, "Deep learning for algorithmic trading: enhancing MACD strategy," in *Proceedings of the 6th Int. Conf. on Comptg. and Artificial Intelligence*, pp. 51–57, Tianjin, China, April 2020.

[12] F. Shah, J. P. Li, F. Shah, and Y. Shah, "Hadoop with Wavelet support for medical big data," in *Proceedings of the 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 17–19 Dec. 2021.

[13] M. Eling, D. Nuessl, and J. Staubli, "The impact of artificial intelligence along the insurance value chain and on the insurability of risks," *The Geneva Papers on Risk and Insurance - Issues and Practice*, vol. 46, no. 2, 2021.

[14] S. Yu, Y. Chen, and H. Zaidi, "AVA: A financial service chatbot based on deep bidirectional transformers," *Frontiers in Applied Mathematics and Statistics*, vol. 7, p. 604842, 2021.

[15] D. J. Caleb, J. P. Li, and F. Shah, "DWT based encryption technique for medical images in cloud databases," in *Proceedings of the 13th ICCWAMTIP*, IEEE, Dec 2016.

[16] M. Leo, S. Sharma, and K. Maddulety, "Machine learning in banking risk management: A literature review," *Risks*, vol. 7, no. 1, 2019.

[17] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *Review of Financial Studies*, vol. 33, no. 5, pp. 2233–2273, 2020.

[18] F. Shah, A. Anwar, I. ul haq, H. AlSalman, S. Hussain, and S. Al-Hadhrami, "Artificial intelligence as a service for Immoral content detection and Eradication," *Scientific Programming*, vol. 2022, pp. 1–9, 2022.

[19] A. Anonymous, "A comparative study: Bank loan vs Trade Credit," *A penal data approach Invitation for 9th CUST International Business Research Conference(CIBRC)*, vol. 20, no. 3, pp. 457–480, 2012.

[20] T. Ye and L. Zhang, "Derivatives pricing via machine learning," *Journal of Mathematical Finance*, vol. 9, no. 3, pp. 561–589, 2019.

[21] T. Mitchell, *Machine Learning*, p. 2, McGraw-Hill, New York, 1997.

[22] R. Triepels, H. Daniels, and R. Berndsen, “Monitoring Liquidity management of banks with Recurrent neural networks,” *Computational Economics*, vol. 57, no. 1, pp. 89–112, 2021.

[23] F. M. Yilmaz and O. Arabaci, “Should deep learning models be in high demand, or should They simply be a very hot Topic? A Comprehensive study for exchange rate forecasting,” *Computational Economics*, vol. 57, no. 1, pp. 217–245, 2021.

[24] J. Yoon, “Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest Approach,” *Computational Economics*, vol. 57, no. 1, pp. 247–265, 2021.

[25] R. R. Kline, “Cybernetics automata studies and the dartmouth conference on artificial intelligence,” in *Proceedings of the IEEE Annals of the History of Computing*, IEEE Computer Society, 2011.

[26] UNDP, “Covid-19 –Pakistan Socio-economic impact Assessment and Response plan,” UNDP, 2020.

[27] R. E. Watkins, F. C. Cooke, R. J. Donovan, C. R. MacIntyre, R. Itzwerth, and A. J. Plant, “Influenza pandemic preparedness: motivation for protection among small and medium businesses in Australia,” *BMC Public Health*, vol. 7, no. 1, p. 157, 2007.

[28] K. Gai, M. Qiu, and X. Sun, “A survey on FinTech,” *Journal of Network and Computer Applications*, vol. 103, pp. 262–273, 2018.

[29] V. Gatteschi, F. Lamberti, C. Demartini, C. Pranteda, and V. Santamaria, “To blockchain or not to blockchain: That is the Question,” *IT Professional*, vol. 20, no. 2, pp. 62–74, 2018.

[30] M. Geranio, “Fintech in the exchange industry: potential for disruption?” *Masaryk University Journal of Law and Technology*, vol. 11, no. 2, pp. 245–266, 2017.

[31] H. Gimpel, D. Rau, and M. Röglinger, “Understanding FinTech start-ups - a taxonomy of consumer-oriented service offerings,” *Electronic Markets*, vol. 28, no. 3, pp. 245–264, 2018.

[32] P. Gomber, R. J. Kauffman, C. Parker, and B. W. Weber, “On the fintech revolution: Interpreting the forces of innovation, disruption, and Transformation in financial services,” *Journal of Management Information Systems*, vol. 35, no. 1, pp. 220–265, 2018.

[33] P. Gomber, J.-A. Koch, and M. Siering, “Digital Finance and FinTech: current research and future research directions,” *Journal of Business Economics*, vol. 87, no. 5, pp. 537–580, 2017.

[34] D. Gozman, J. Liebenau, and J. Mangan, “The innovation Mechanisms of fintech start-ups: Insights from SWIFT’s Innotribe competition,” *Journal of Management Information Systems*, vol. 35, no. 1, pp. 145–179, 2018.

[35] J. Gruin and P. Knaack, “Not just Another shadow bank: Chinese Authoritarian capitalism and the ‘developmental’ Promise of digital financial innovation,” *New Political Economy*, vol. 25, no. 3, pp. 370–387, 2020.

[36] E. Hernández, M. Öztürk, I. Sittón, and S. Rodríguez, “Data protection on fintech Platforms Communications in Computer and Information Science,” in *Proceedings of the International Conference on Practical Applications of Agents and Multi-Agent Systems*, pp. 223–233, Springer, Cham, 2019.

[37] V. P. Mishra and B. Shukla, “Development of simulator for intrusion detection system to detect and alarm the DDoS attacks,” in *Proceedings of the 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)*, pp. 803–806, IEEE, 2018.

[38] W. Meng, L. Zhu, W. Li, J. Han, and Y. Li, “Enhancing the security of FinTech applications with map-based graphical password authentication,” *Future Generation Computer Systems*, vol. 101, pp. 1018–1027, 2019.

[39] L. Bo and X. Zeshui, “Insights into financial technology (FinTech): a bibliometric and visual study,” *Financial Innovation*, vol. 7, 2021.

[40] E. Z. Milian, M. d. M. Spinola, and M. M. d. Carvalho, “Fintechs: A literature review and research agenda,” *Electronic Commerce Research and Applications*, vol. 34, p. 100833, 2019.

[41] D. Tabitha and G. Stella, “Digital finance and its impact on financial inclusion,” *Journal of Emerging Technologies and Innovative Research*, vol. 6, no. 1, 2019.

[42] K. H. Hsieh and N. P. D. Candidates, *Open Banking: A Case Study for Open API Economy*.

[43] L. Brodsky and L. Oakes, *Data sharing and Open Banking*, McKinsey & Company, 2017.

[44] D. L. K. Chuen and R. H. Deng, *Handbook of Blockchain, Digital Finance, and Inclusion: Cryptocurrency, FinTech, InsurTech, Regulation, ChinaTech, mobile security, and Distributed Ledger*, Academic Press, 2017.

[45] D. W. Arner, J. Barberis, and R. P. Buckley, *FinTech and RegTech in a Nutshell, and the Future in a Sandbox*, CFA Institute Research Foundation, 2017.

[46] N. G. Packin, “RegTech, compliance and technology judgment rule,” *Chicago Kent Law Review*, vol. 93, p. 193, 2018.

[47] M. L. Fein, “Robo-advisors: A closer look,” Available at SSRN 2658701, 2015.

[48] D. Lee and R. H. Deng, *Handbook of Blockchain, Digital Finance, and Inclusion: Cryptocurrency, FinTech, InsurTech, and Regulation*, Academic Press, 2018.

[49] M. Kiri, “Effect of fintech strategy on financial services delivery to the unbanked low-income earners in Nairobi county,” (doctoral dissertation, school of business, university of Nairobi), 2018.

[50] T. H. Lee and H. W. Kim, “An exploratory study on FinTech industry in Korea: crowdfunding case,” in *Proceedings of the 2nd International conference on innovative engineering technologies (ICIET’2015)*Bangkok, 2015.

[51] C. W. Cai, “Disruption of financial intermediation by FinTech: a review on crowdfunding and blockchain,” *Accounting and Finance*, vol. 58, no. 4, pp. 965–992, 2018.

[52] J. Kang, “Mobile payment in FinTech environment: trends, security challenges, and services,” *Human-centric Computing and Information Sciences*, vol. 8, no. 1, p. 32, 2018.

[53] B. Li, S. D. Hanna, and K. T. Kim, “Who Uses mobile payments: fintech potential in users and non-Users,” *Journal of Financial Counseling and Planning*, vol. 31, no. 1, pp. 83–100, 2020.

[54] L. Lu, “Decoding Alipay: mobile payments, a cashless society and regulatory challenges,” *Butterworths Journal of International Banking and Financial Law*, vol. 33, no. 1, pp. 40–43, 2018.

[55] T. Philippon, *The FinTech Opportunity*, National Bureau of Economic Research, 2016.

[56] J. Lerner and P. Tufano, “The consequences of financial innovation: a counterfactual research agenda,” *Annual Review of Financial Economics*, vol. 3, no. 1, 2011.

[57] I. Goldstein, W. Jiang, and G. A. Karolyi, “To FinTech and beyond,” *Review of Financial Studies*, vol. 32, no. 5, pp. 1647–1661, 2019.

[58] P. Schueffel, “Taming the beast: a scientific definition of FinTech,” *Journal of Innovation Management*, vol. 4, no. 4, pp. 32–54, 2016.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9829570, 1 page
https://doi.org/10.1155/2023/9829570

Retraction

Retracted: Application Research of Particle Swarm Algorithm in Bank Human Resource Management

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wu, S. Wang, X. Wang, and Z. Wang, "Application Research of Particle Swarm Algorithm in Bank Human Resource Management," *Security and Communication Networks*, vol. 2022, Article ID 8788894, 14 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8788894, 14 pages
https://doi.org/10.1155/2022/8788894

Research Article

Application Research of Particle Swarm Algorithm in Bank Human Resource Management

You Wu[1], **Shengqi Wang**[2], **Xing Wang**[1], **and Zheng Wang**[1]

[1]*School of Public Administration, Guangzhou Xinhua University, Guangzhou 13902, China*
[2]*China Construction Bank Corporation Zhuhai, School of Business, Guilin University of Electronic Technology, Guilin 10595, China*

Correspondence should be addressed to Zheng Wang; wzwjw@xhsysu.edu.cn

Received 7 January 2022; Revised 29 January 2022; Accepted 4 February 2022; Published 7 March 2022

Academic Editor: Chin-Ling Chen

In the era of knowledge economy, human resources as the first resource of enterprises have long become a consensus. However, human resource management in my country is still in its infancy. There are few studies on the relationship between the two, and the research conclusions are inconsistent. There are very few subjects for research. As the core of my country is banking industry, state-owned commercial banks affect the development direction of my country's banking industry, and their human resource management has its own unique features. However, due to my country's special national conditions, state-owned commercial banks have not been fully market-oriented, lacking a certain degree of independence and autonomy, and the impact of administrative intervention has led to human resource management in state-owned commercial banks that have not been developed as they should be. In a series of questions, in today's turbulent global financial environment, state-owned commercial banks need to improve their human resource management to enhance their sustainable competitiveness. Based on combing the research on the relationship between human resource management and corporate performance by domestic and foreign scholars and analyzing the status quo and problems of human resource management of state-owned commercial banks, this paper collects data through questionnaires and analyzes the reliability and validity of the data. Then use empirical analysis methods to study the relationship between human resource management of state-owned commercial banks and corporate performance. The research results show that there is a significant positive correlation between the human resource management practices of state-owned commercial banks and corporate performance, that is, improving the level of human resource management can promote the improvement of corporate performance.

1. Introduction

At the Third Session of the Twelfth National People's Congress, Zhang Jiangsu, the governor of China Construction Bank, once said "banks are a disadvantaged group." As soon as this remark came out, many people who did not know it just took it as a joke. In fact, Zhang Jiangsu believes that there are huge challenges in the development of the current banking industry: first, depositors demand high returns, and banks raise the interest rate on deposits to absorb deposits, which indirectly raises the financing cost of the whole society; second, lenders criticize banks for not only making loans difficult but also the loans are expensive. From this perspective, banks are a disadvantaged group. In addition, judging from the current fiercely competitive financial environment, the current situation of the banking industry is said to have both internal and external troubles [1–3]. Internally, financial marketization reform has already begun. As we all know, the core competitive advantage of enterprises lies in talents, especially for the knowledge-intensive banking industry. The opening of the upper limit on deposit interest rates is a "step at the door," and the market-oriented reform of bank interest rates announced as completed. Banks are facing increasing pressure on deposits, and the battle for deposits among banks will inevitably intensify. Foreign troubles should also not be underestimated. Since 2001, China Joining the WTO until 2006, the five-year protection period has expired, and the banking industry fully opened on December 11, 2006 [2, 4–6]. On the one hand, state-owned commercial banks are facing fierce competition from overseas banks; on the other hand, they are facing the encroachment of Internet finance. From the recent two years of Aliped and Yu'ebao to the wave of electronic red envelopes during the Spring Festival Gala this year, Internet companies have set off a new round of financial wars. People cannot help asking, why are emerging Internet companies standing in front of Internet finance. Human resources are the sum of knowledge and skills condensed in employees and have strong background dependencies. The management of human resources is difficult to be imitated or copied by competitors, which is where the enterprise obtains a sustainable competitive advantage. The author believes that because my country's state-owned commercial banks have not been fully market oriented, lacking independence, and autonomy, they cannot sharply capture market information and carry out reforms and innovations in a timely manner [7–10].

In the end, private finance has already shown its footing in the competition with state-owned commercial banks. With the continuous decline of my country's banking market access infrastructure, as of the end of 2014, there were more than 4,000 registered legal entities in the domestic banking industry, and various financial and nonfinancial institutions began to expand related businesses to the banking industry, hoping to win a share in the financial industry soup [11–13]. According to the 2014 Bank of China Annual Report, with the acceleration of interest rate marketization, deposit substitute products have become more abundant, pressure on bank deposits to drain has increased, and deposit growth has declined. With the slowdown of domestic economic growth, the traditional profit model of commercial banks relying solely on asset expansion is changing. Under the more complex and challenging environment of the bank's operating environment, the past "three highs" development model of my country's banking industry (high scale growth, high-performance indicators, and high profit growth) is no longer feasible. It only hoped to rely on widening interest margins [5, 14, 15]. The traditional profit model of increasing credit assets is no longer a long-term solution. In addition, the capital market is increasingly competing for the banking market. The state-owned commercial banks need to improve their own risk management and control capabilities and the ability to respond to market risks for the slowdown of domestic economic growth. To maintain or improve their competitive position, state-owned commercial banks must pay attention to their own human resources. Enterprises should improve their human resource management level. To sum up, it can be seen that the importance of state-owned commercial banks in the financial system is getting lower and lower. State-owned commercial banks are only strong on the outside and are currently facing a huge challenge to survival and development [16–18].

As we all know, the core competitive advantage of enterprises lies in talents, especially for the knowledge-intensive banking industry. Human resources, which is the sum of knowledge and skills condensed in employees, have a strong background dependence and are difficult to be imitated or plagiarized by competitors, making it a place where companies can obtain sustainable competitive advantages. State-owned: to maintain or improve its competitive position, commercial banks must attach importance to their own human resources and improve their own human resource management level. At present, for most banks, especially state-owned commercial banks, the human resource management of enterprises is still in the traditional personnel management stage, and there are problems such as weak human resource management concepts, backward management levels, and imperfect system construction [19–21]. In the end, talents may appear. Disadvantages that hinder the development of the enterprise are loss, technological disconnection, and overall aging of employees. In the process of market competition, those human resource management models that are out of touch with corporate development strategies, unscientific recruitment and allocation, and imperfect training and incentive mechanisms have been difficult to adapt to the needs of corporate competition and development requirements. Sound development is bound to have a negative impact on corporate organizational performance. Therefore, it is necessary to make further research on the relationship between human resource management of state-owned commercial banks and corporate performance [22–24].

This paper uses People's Republic of China's state-owned commercial banks as sample data to conduct empirical correlation and regression analysis to study the impact of state-owned commercial banks' human resource management practices on corporate performance. Through the research, it is expected to achieve the following goals [22, 25, 26]. First, through the investigation of the human resource management level and corporate performance of our state-owned commercial banks, analyze whether the impact of human resource management practices on corporate performance exists in the state-owned commercial banks. Second, based on the quantitative empirical research results, certain countermeasures and suggestions put forward for the improvement of the human resource management level of state-owned commercial banks. Scholars at home and abroad have done a lot of research on the relationship between human resource management and corporate performance, but Chinese scholars started late in the research on the relationship between the two, the research objects are only individual companies or individual industries, and the research conclusions are also inconsistent. As a result, many companies have insufficient understanding of the relationship between the two and cannot correctly understand the value of human resource management [6, 27, 28]. The research logical structure of this paper is shown in Figure 1.

2. Existing Research Results and Literature Review

By reviewing the development process of human resource management, we can see that human resource management can be divided into three stages in general. First, it is the personnel management stage that focuses on personnel

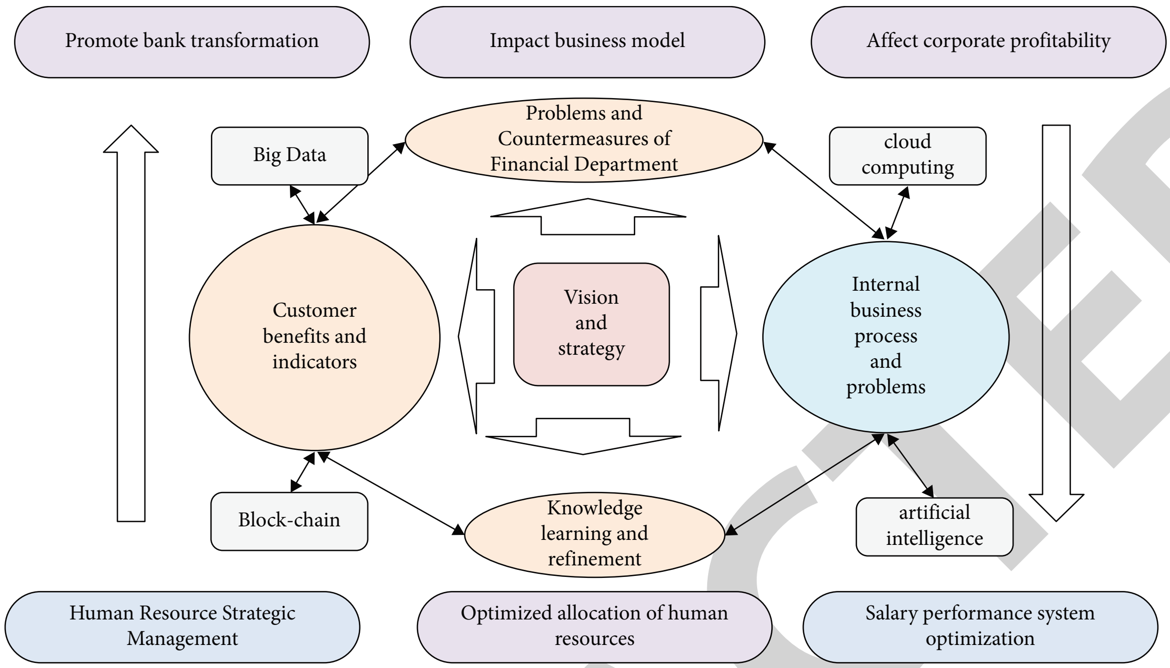


Figure 1: The research logical structure of this paper.

work; second, the human resource management stage that focuses on the role of people; and finally, the human resource management raised to the strategic level of strategic human resource management.

2.1. Research on Human Resource Management Theory. Robert Irving, the famous British utopian socialist in the early 19th century, conducts experiments in the factories it operates. The experiments include methods such as improving working conditions, shortening working days, and increasing wages, in order to explore methods and systems that are beneficial to workers and factories. Therefore, he is called "the father of personnel management." The stage of scientific management at the end of the 19th century and the beginning of the 20th century is represented by American management scientist Taylor. He began to study how to improve the productivity of individual workers and the overall efficiency of the enterprise. In the current state, for most banks, the human resource management of enterprises is still in the traditional personnel management stage. In particular, the state-owned commercial banks have problems such as weak human resource management concept and backward management level. Taylor's contribution is that he put forward ideas such as employee work quotas, work process standardization, employees' ability to adapt to their jobs, and differentiated piecework for salary settlement. At that time, it caused people's attention and led to the development of human resource management thought [14, 26, 29].

In 1929, the American psychologist and management scientist Mayo conducted a nine-year Hawthorne experiment at the Hawthorne factory of Western Electric Company and began to study the behavior of people in the organization. He believes that workers are social people rather than economic people, there are informal organizations in enterprises, and production efficiency mainly depends on the working attitude of workers and the relationship with people around them. Subsequently, behavioral sciences developed vigorously in the United States, resulting in a large number of influential behavioral scientists and their theories, including Maslow's theory of needs, McCree go's XY theory, Herzberg's two-factor theory, Rome's theory, and expectation theory. Model analysis results are shown in Figure 2.

2.2. Corporate Performance Research. In the 1980s, Milles and others began to conduct systematic research on human resource management. They believed that traditional personnel management was no longer suitable for the needs of modern enterprises and put forward the concept of human resource management. This kind of change is not a simple name change, but a huge change in management philosophy and management methods. Regarding corporate employees as the most important resource for achieving corporate strategic goals, they have flexibility, creativity, and plasticity that other resources do not have and attach importance to the long-term development of employees. Second, they believe that human resource management needs to participate in the formulation and implementation of corporate strategic decisions. The goal of human resource management is not only to provide enterprises with the required human resource products and services but also to improve their operating profits, survival, and development capabilities and reduce labor costs. Increasing the competitiveness of enterprises, etc., also pay more attention to the role of human resources. Regarding the content of human resource management,

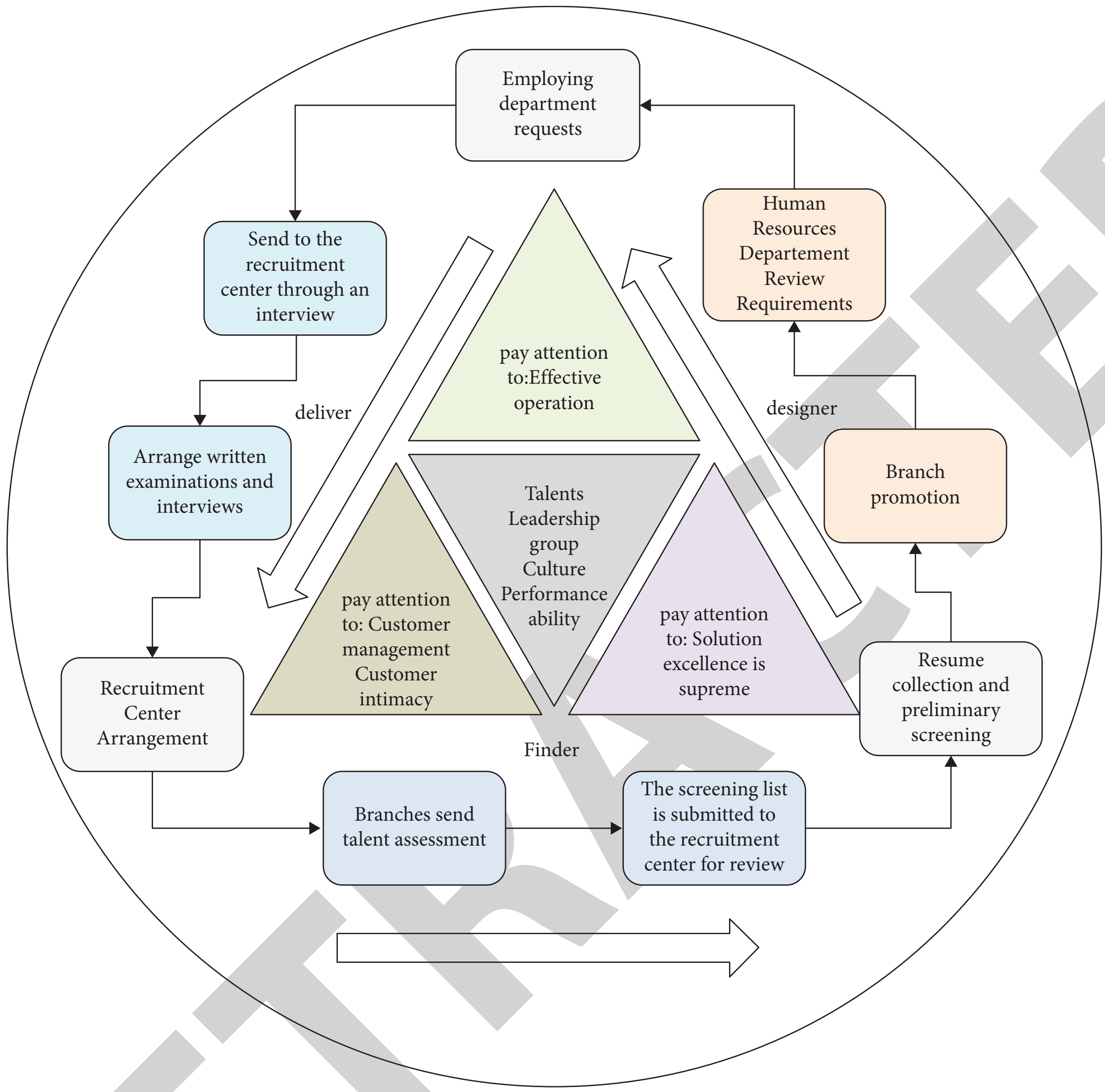


Figure 2: The relationship diagram of the influence of financial technology on the human resource management of state-owned enterprise banks.

scholars at home and abroad have different opinions and give their own rationale from different angles [30].

Academia generally divides human resource management into six major modules: (1) human resource planning: refers to forecasting the supply and demand of human resources in the enterprise according to changes in the internal and external environment and conditions of the enterprise. They needs to be balanced to ensure the smooth progress of human resource management practices from the source. The construction of the state-owned banking system is not perfect. These problems are time honored and can lead to brain drain, technological disconnect, and an overall aging workforce. Human resource planning is a decisive plan in enterprise planning, and it has a guiding role in human resource management activities. It is of strategic significance and called the link of human resource management activities. (2) Recruitment and allocation of personnel: whether the recruitment is appropriate not only relates to whether the work can carried out smoothly but also determines whether the enterprise can develop rapidly. Therefore, it is necessary to choose different recruitment channels and recruitment methods according to the characteristics of the enterprise and the characteristics of the position. In terms of staff allocation, it is necessary to make the best use of their talents and match their positions to adapt their personal ability level to the job requirements. The interrelationships between model elements are shown in Figure 3.

2.3. Research on the Relationship between Human Resource Management and Enterprise Performance. The understanding of corporate performance by domestic and foreign scholars has roughly gone from focusing on results to focusing on process, from financial performance to nonfinancial performance. This article reviews the connotation of corporate performance from a multidimensional perspective

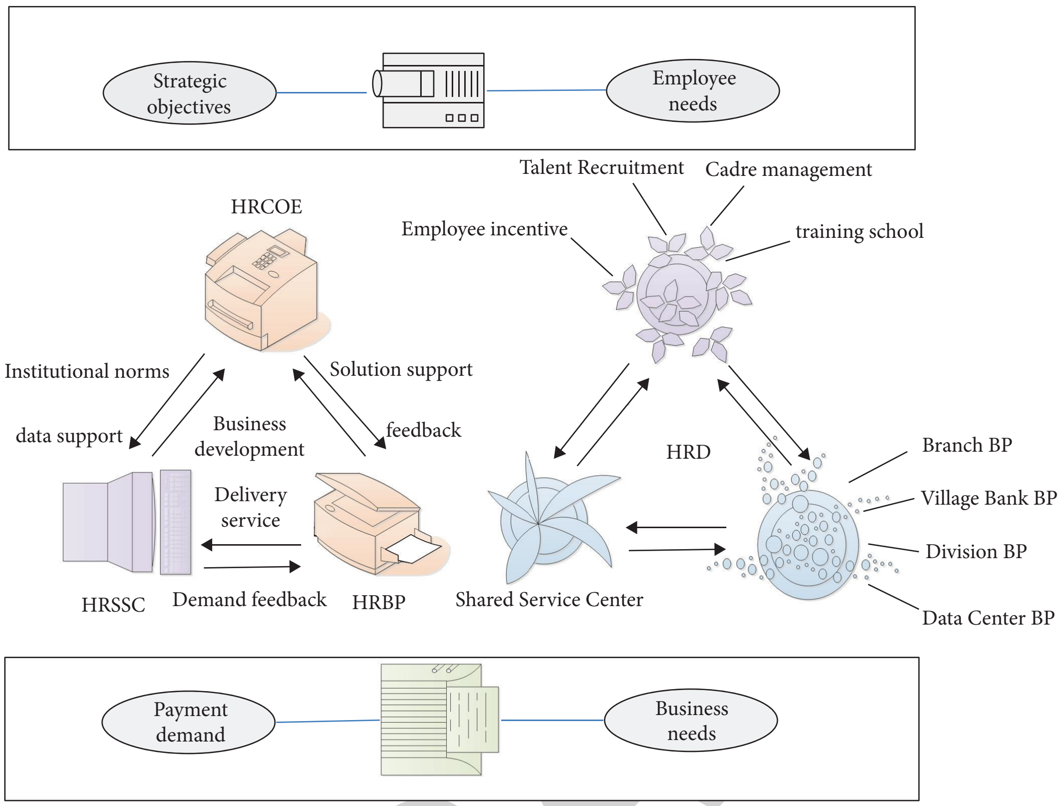


Figure 3: Schematic diagram of the ecological allocation of talent resources for postgraduate joint training.

and focuses on companies that take into account both financial performance and nonfinancial performance. The performance measurement methods, which from the perspective of influencing factors, are the survival and development of enterprises in society. It is affected by different internal and external factors, and the results of corporate performance are also affected by various factors. The stakeholder management theory put forward by Ferryman in the book "Strategic Management: The Analysis Method of Stakeholder Management" believes that the development of any enterprise cannot be separated from the input and participation of various stakeholders. The overall interests of all stakeholders: while pursuing good performance, enterprises must also consider the attitudes of internal and external stakeholders. Stakeholders generally include shareholders. Shareholders are investors. Enterprises must first consider maximizing the interests of shareholders. The second is customers. Customers hope to obtain the most cost-effective services and products in the enterprise. The third is corporate employees. Employees hope to get rewards consistent with their efforts in the company. The fourth is the community. They hope that companies can bring development to the local economy while fulfilling corporate social responsibilities and paying attention to protecting the ecological environment.

Corporate performance includes three levels, namely organizational performance, group performance, and individual employee performance. This article is positioned on the organizational performance level. Organizational performance is the overall performance of the organization, which is generally considered by the operating efficiency and the performance of the operators during a certain operating period of the organization. The level of corporate operating efficiency is mainly manifested in the profitability, asset operation level, debt solvency, and subsequent development capabilities; the performance of the operator is mainly based on the results and achievements of the operator's operation, growth, and development in the process of operating and managing the enterprise. The contribution made and other aspects are measured. There are many traditional measurement methods of organizational performance, but they mainly focus on financial indicators. This article focuses on the measurement method proposed by Kaplan and Norton in "Balanced Scorecard" in 1992. Therefore, the human resource management reform of state-owned commercial banks: the reform of state-owned enterprises needs to do a good job in human resources planning from the source and provide excellent human resources for the survival and development of enterprises in a timely manner. They believe that in addition to financial indicators, companies must also pay attention to nonfinancial indicators. Financial indicators include four dimensions. The four dimensions influence and promote each other, and the central idea is to realize the company's vision and mission.

3. Analysis on the Current Situation of Human Resources Management in State-Owned Commercial Banks

Through the above analysis, it seen that there are many problems in the actual operation of the human resource management of state-owned commercial banks. Through empirical analysis and research, it verified that the six parts of human resource management practice indeed have a significant impact on the corporate performance of state-owned commercial banks. State-owned commercial banks need to improve their own human resource management level to improve corporate performance. This article combines the theoretical knowledge of human resource management to put forward some concrete countermeasures and suggestions for the reform of human resource management of state-owned commercial banks from the perspective of human resource management.

3.1. Improve Recruitment Management and Optimize Personnel Structure. The previous article mentioned that human resource planning has guiding significance for enterprise human resource management activities. Therefore, the human resource management reform of state-owned commercial banks can first do a good job of human resource planning from the source and provide excellent human resources for the survival and development of enterprises in a timely manner. In this research, the ability of human resource planning significantly and positively correlated with corporate performance, which shows that the development of a sound human resource plan can significantly improve corporate performance and gain competitive advantage. Therefore, my country's state-owned commercial banks should integrate human resource planning into their own reform and development strategies and truly establish a "people-oriented" strategic human resource management concept. State-owned commercial banks need to do the following in detail. Model analysis results are shown in Figure 4.

3.1.1. Do a Good Job Analysis. The main task of job analysis is to formulate clear job specifications and work instructions for each position in the company, to help employees quickly master work skills and provide the best direction for employee recruitment. Human resource planning ability has a significant positive correlation with corporate performance. This phenomenon shows that the development of a sound human resource planning can significantly improve the performance of the enterprise and obtain a competitive advantage. Therefore, it is more necessary work instructions and job specifications that meet the characteristics of the job must made for each job. The human resources department can formulate the job description of the job by understanding the job content, job requirements, rights, responsibilities and benefits of each job, and the quality requirements of each job.

3.1.2. Formulate Perfect Staffing Standards. Reasonable labor quota is the scientific standard for enterprise employment, ensuring that enterprises use human resources rationally and economically, and improve labor productivity. The long-standing problems of redundant staff and inbreeding in state-owned commercial banks can establish by setting clear staffing standards to remove noncompliant employees, reduce labor costs for enterprises, and improve the efficiency of talent utilization. Recruitment is not only an issue of the number of employees but also the quality of human resources. Therefore, state-owned commercial banks must introduce high-quality talents to make the best use of their talents and make them suitable. In addition, with changes in the internal and external environments of enterprises, state-owned commercial banks should make appropriate amendments to the staffing standards that do not meet actual needs. Model analysis results are shown in Figure 5.

3.1.3. Institutionalization of Human Resource Management. The most important thing in human resource planning is to establish and improve various personnel rules. Therefore, my country's state-owned commercial banks should integrate human resource planning into the enterprise's own reform and development strategy. State-owned enterprises should truly establish a "people-oriented" strategic human resource management concept.

The calculation principle of local consistency is relatively simple, mainly using Kendal Concorde coefficient, and the specific calculation formula is as follows:

$$W = \frac{\sum (R_i)^2 - n(\overline{R})^2}{1/12\mathrm{K}^2\left(n^3 - n\right)}. \tag{1}$$

The calculation formula of single-sample statistics is as follows:

$$t = \frac{\overline{X} - \mu}{\delta_x/\sqrt{n-1}}. \tag{2}$$

Set two random sequences X and Y, Pearson correlation coefficient between the two sequences is r, then:

$$\begin{aligned} r &= \frac{\operatorname{cov}(X, Y)}{\sqrt{\sigma_x^2}\sqrt{\sigma_y^2}} \\ &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}}. \end{aligned} \tag{3}$$

In order to ensure the accuracy of the results, this paper uses two evaluation indexes, mean absolute error, and root mean square error, to evaluate the optimization effect of the model. The specific calculation formulas are as follows:

$$\mathrm{MAE} = \frac{1}{s}\sum_{i=1}^{s}\left|\hat{y}_i - y_i\right|, \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{s}\sum_{i=1}^{s}\left[\hat{y}_i - y_i\right]^2}. \tag{5}$$

Coverage index calculates the ratio of predicted items to all unscored items, so as to measure the comprehensiveness of prediction. Assuming that h items are predicted, the calculation method of coverage is as follows:

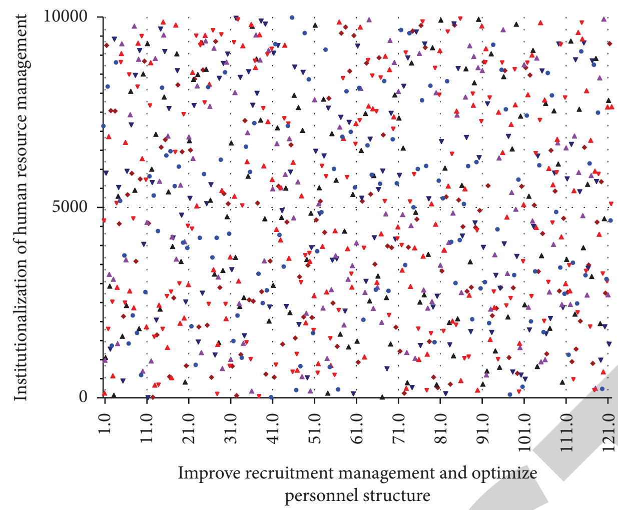


Figure 4: The applicability of developing a sound human resource plan.

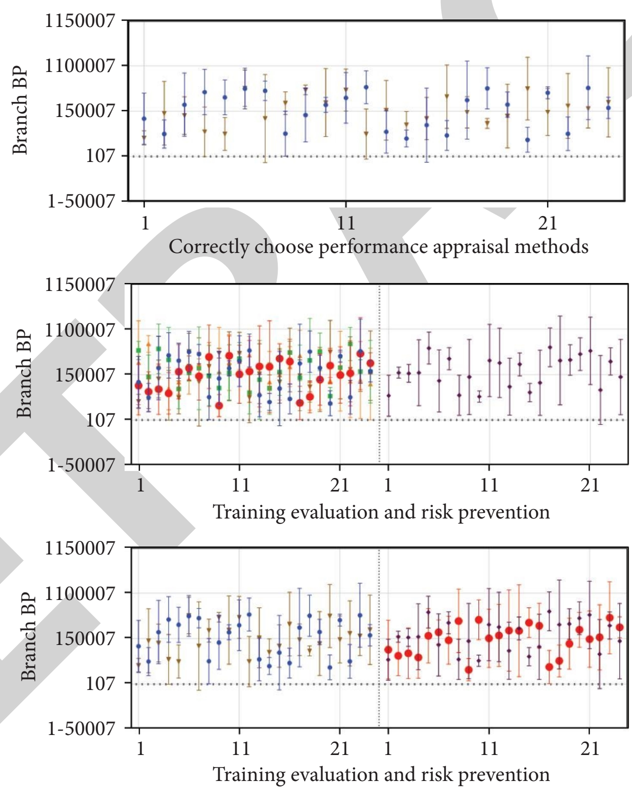


Figure 5: Model of improve recruitment management and optimize personnel structure.

$$\mathrm{Cov} = \frac{h}{n}. \tag{6}$$

Recall index is also used to evaluate the system effect in the field of information retrieval. The larger the recall value, the better is the recommendation quality of the algorithm.

$$\begin{aligned}\mathrm{recall} &= \frac{\mathrm{Hits}}{|\mathrm{test}|} \\ &= \frac{|\mathrm{test}\ \mathrm{I}\ \ \mathrm{Top} - N|}{|\mathrm{test}|}.\end{aligned} \tag{7}$$

On the one hand, it helps to weaken the long-existing "official standard" idea of my country's state-owned commercial banks, realize the transparency and democratization of internal management, and alleviate employee conflicts. On the other hand, the process of institutionalization can solicit employees' opinions and reflect employee awareness. Only by involving employees in formulating systems can make the employees better comply. Model analysis results are shown in Figure 6.

3.2. Clarify the Criteria for Talent Selection and Evaluate the Effectiveness of Recruitment. The role of recruitment management for talents related to the question of whether state-owned commercial banks can improve the quality of human resources. The impact of recruitment management on corporate performance has also confirmed. State-owned commercial banks need to pay attention to the following points in improving recruitment management. The effect of recruitment and selection affects whether the human resource demand of state-owned commercial banks is realized. The level of talent demand of state-owned commercial banks determines that the form of recruiting talents cannot be single. For different talent needs, different forms of recruitment should be adopted. Model analysis results are shown in Figure 7.

3.2.1. Internal Recruitment. When there are vacancies in the company, you can first consider adopting the internal recruitment method to select suitable personnel from the company's internal personnel to supplement the vacant positions. Because managers have a fuller understanding of internal employees, employees are also more familiar with the job content and can adapt to new jobs more quickly. Therefore, the accuracy of internal recruitment is high. In addition, internal recruitment can provide employees with development opportunities and generate certain incentives. State-owned commercial banks can use the form of internal recruitment to recommend outstanding talents through employee recommendation, and they also publish recruitment information within the company to allow employees to compete for posts and form an enterprising spirit within the company.

In order to alleviate the impact of differences in actual scores between users on user trust relationship [28], the mean value of the scoring difference between the two users on the common scoring items is calculated as follows:

$$\varepsilon = \frac{\sum_{i \in I_a \mathrm{I}\, I}\left(\left|c_{ai} - c_{bi}\right|\right)}{\left|I_a \mathrm{I}\, I_b\right|}. \tag{8}$$

The constraint conditions are equations (3)–(5):

$$\begin{aligned} f\left(x_i, \omega\right) - y_i &\le \xi_i + \square, \quad i = 1, 2, \ldots l, \\ \xi_i \xi_j &\ge 0, \quad i = 1, 2, \ldots l. \end{aligned} \tag{9}$$

Shape the general functional relationship between the output y of the injury model and the input $x_1, x_2, \ldots, x_n$. The Kolmogorov–Gabor polynomial is as follows:

$$\begin{aligned} y &= f\left(x_1, x_2\right) \\ &= a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2. \end{aligned} \tag{10}$$

And treating each of the monomials as m input models in the original structure of the modeling network:

$$\begin{aligned} v_1 &= a_0, \\ v_2 &= a_1 x_1, \\ v_3 &= a_2 x_2, \ldots, \\ v_6 &= a_5 x_1 x_2. \end{aligned} \tag{11}$$

The final information $i_t \times C_t'$ is expressed as the value that can be obtained C_t from the output information of the joint forgetting gate:

$$C_t = f_t * C_{t-1} + i_t * C_t'. \tag{12}$$

3.2.2. External Recruitment. Compared with internal recruitment, externally recruited talents can bring new ideas and new methods of management to the company, and produce a "surplus fish effect" to stimulate employees' morale. The main ways of external recruitment is to publish advertisements. The other is to use intermediaries to recruit the required senior and cutting-edge talents through headhunting companies. The main task of job analysis is to formulate clear job specifications and job descriptions for each job in the enterprise. The analysis results are helpful for employees to quickly master work skills and also provide the best direction for employee recruitment. The third is commonly used campus recruitment, and the fourth is online recruitment. The cost is low, and it not restricted by time and place. The fifth is to recommend talents through acquaintances, with a certain degree of reliability. Model analysis results are shown in Figure 8.

Both internal and external recruitment methods have their own advantages and disadvantages. State-owned commercial banks should choose suitable recruitment methods in accordance with the requirements of the vacant positions for talents. Although campus recruitment can reshape talents that meet the characteristics of the enterprise, college students often have the problem of being superior and inferior, inaccurate in positioning themselves, and confused about the future development, and it is difficult to determine their suitable jobs and positions. With regard to psychological expectations, there may be a higher turnover rate. Therefore, the state-owned

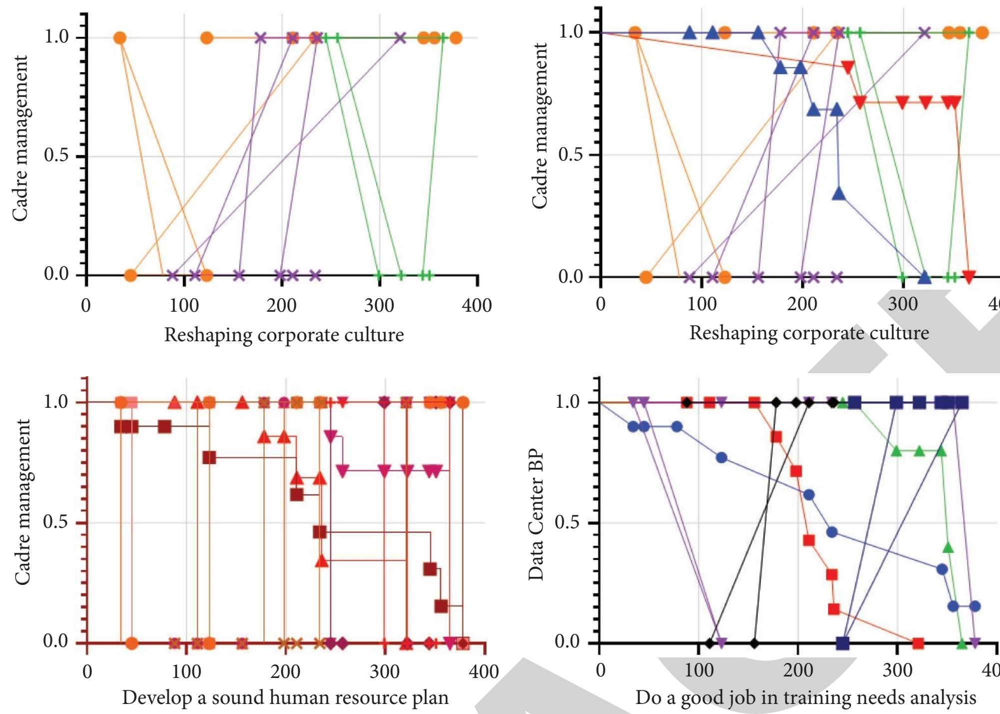


FIGURE 6: New public management theory in choosing the right recruitment method.

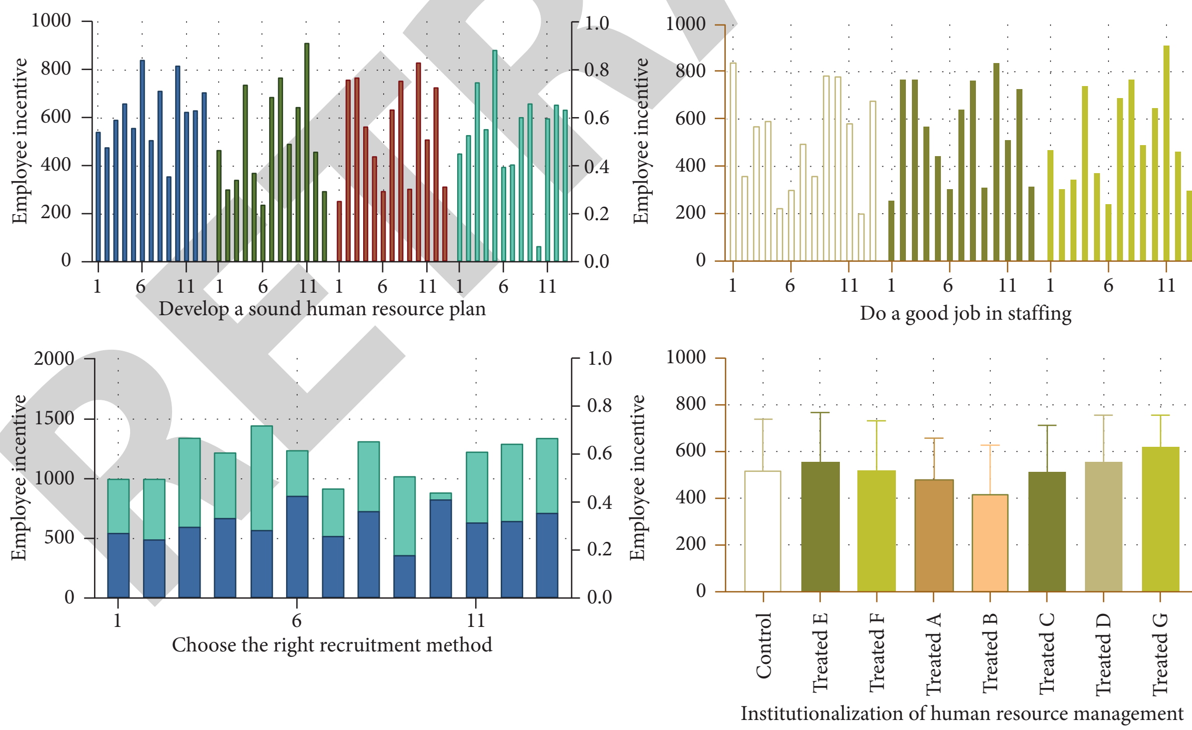


FIGURE 7: System analysis results after employee training.

commercial banks should pay more attention to their talents with the positions It also helps them determine the suitable positions. State-owned commercial banks should choose suitable recruitment methods in accordance with the requirements of the vacant positions for identifying talents. Model analysis results are shown in Figure 9.

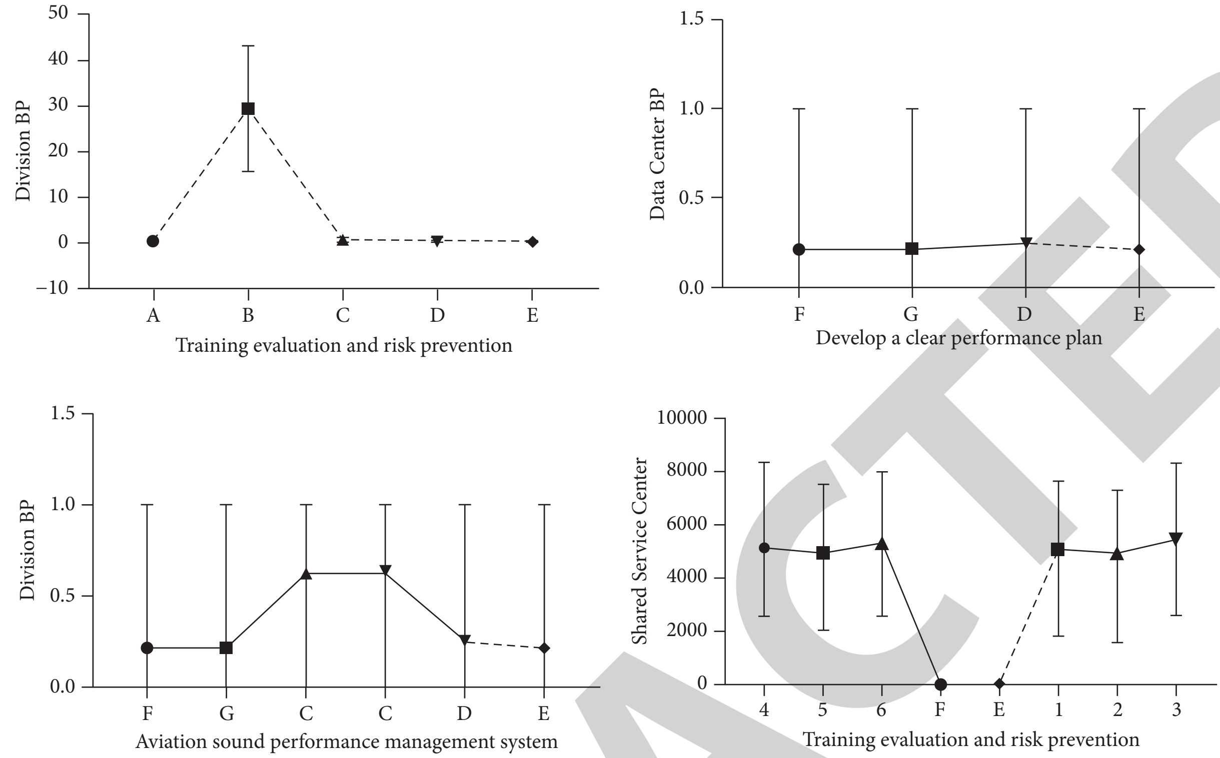


FIGURE 8: Training evaluation and risk prevention and choice of training method.

3.3. Aviation Sound Performance Management System

3.3.1. Clarify the Criteria for Talent Selection and do a Good Job Evaluation on Recruitment Effect. State-owned commercial banks can only find talents that meet the job requirements more quickly and accurately after clarifying the criteria for talent selection. The talent selection criteria should clearly state in the job analysis, that is, there must be clear employee appointment standards, including the requirements for the employees' academic qualifications, skills, and physical health. After the recruitment is completed, the effect of each recruitment should be evaluated, the success or failure of this recruitment should be summarized as a reference for future recruitment work, and lessons should be learned to prevent the same mistakes in the next recruitment work. In addition, the recruitment assessment can analyze whether the talents needed recruited and what kind of recruitment needs in the next step. Model analysis results are shown in Figure 10.

3.3.2. Make a Good Job in Staffing. Scientific staffing should select suitable positions according to the characteristics of employees. However, in practice, companies often ignore the characteristics of employees and only arrange personnel work according to the needs of the company, and there is a phenomenon of mismatch between personnel and posts. In the staffing of state-owned commercial banks, new hires can rotate first, so that each new hire can work in each position for a period, so that employees can find out which position they are more suitable. Training and inspections help employees choose positions that are more suitable for their development. The internship period of new personnel can adopt the traditional form of masters and apprentices. On the one hand, it is helpful for new personnel to integrate into the interpersonal relationship of the enterprise more quickly; on the other hand, it allows employees to adapt to work requirements and master work skills more quickly. Model analysis results are shown in Figure 11.

The diversity of training needs of state-owned commercial banks determines the importance of training needs analysis. If there is no training needs analysis, it will only make the training work in a headless state. When employees' knowledge and skills cannot meet the development needs of the company, individual employees may not take the initiative to ask the company to provide training. At this time, the company needs to analyze the training needs of internal employees like what type of training is required and which employees needed training. Carrying out training, trainer selection, training materials, training time and location, and other information are needed to develop a complete training plan.

State-owned commercial banks have many jobs, and each job has multiple levels. Therefore, it is more necessary to make job descriptions and job specifications conform to the job characteristics for each position. Reasonable labor quota is the scientific standard for enterprise employment. Enterprises should ensure rational and economical use of human resources and improve labor productivity. To recognize the importance of human resources and human resource management, we find a human resource management model that suits its own characteristics as soon as

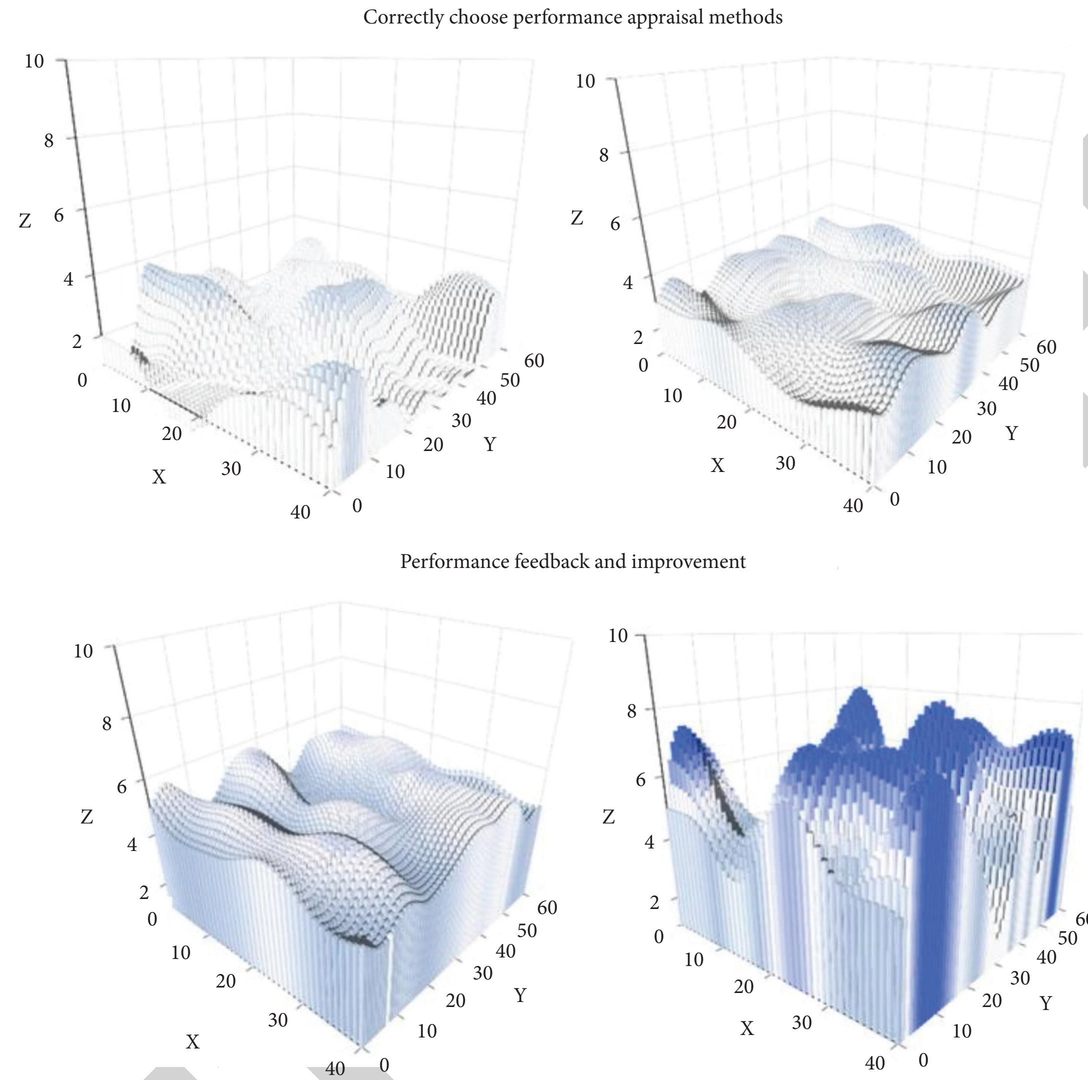


Figure 9: Correctly choose performance appraisal methods and performance feedback and improvement.

possible and improve the level of human resource management to improve its competitive advantage toward domestic and foreign peers. It provides more dimensional thinking for the research of human resource management in our country. Although the existing literature has a lot of research on the relationship between human resource management and corporate performance, there is also a lot of research on the human resource management of state-owned commercial banks. However, no scholar has combined the two to study the impact of state-owned commercial banks' human resource management on corporate performance. By verifying the relationship between human resource management of state-owned commercial banks and corporate performance, this paper believes that there is a significant positive correlation between the two and analyzes the degree of influence of different human resource management functions on corporate performance.

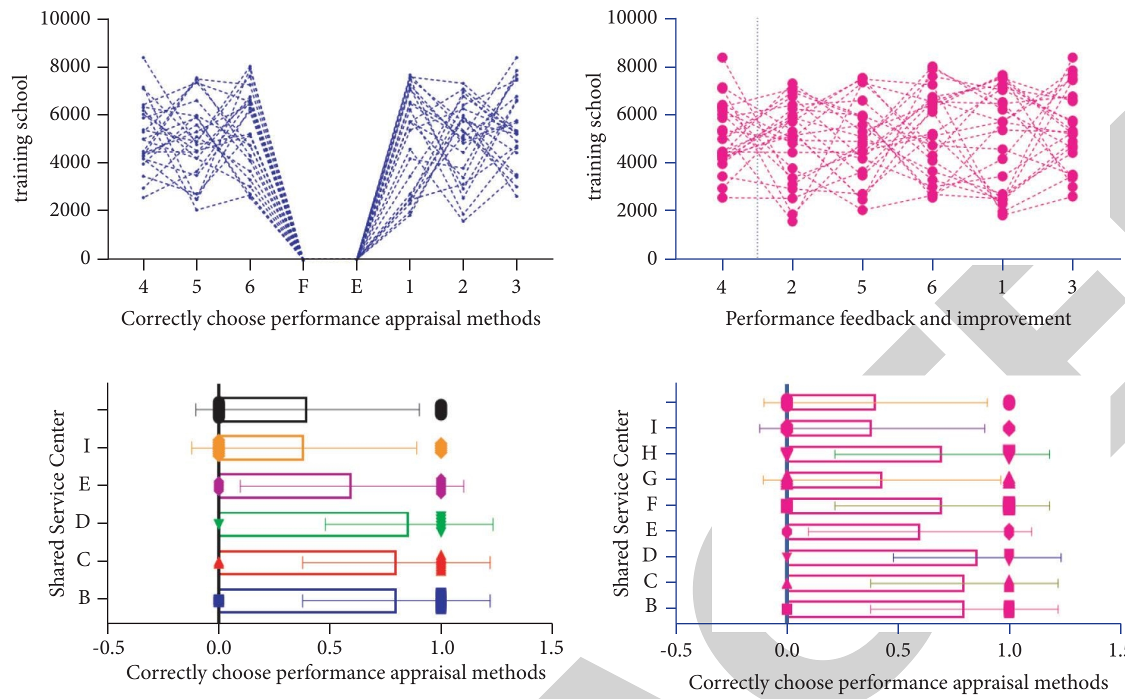


Figure 10: The improvement of salary management system and reshaping corporate culture.

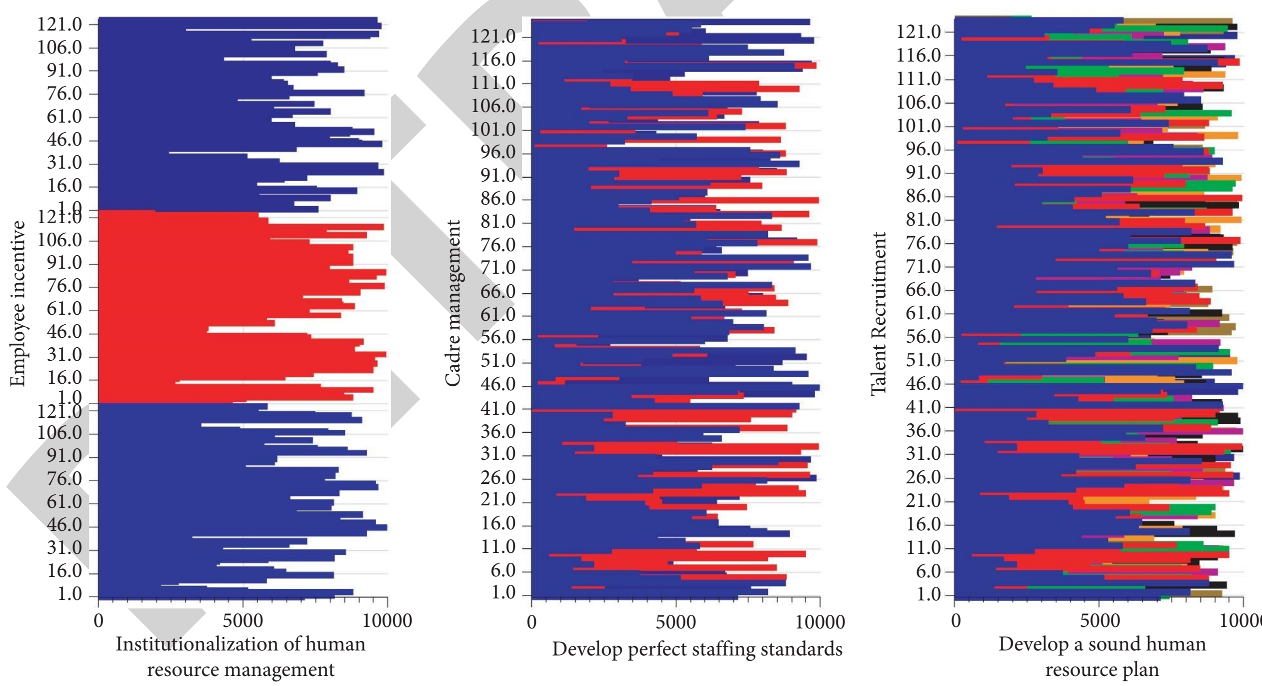


Figure 11: Institutionalization of human resource management and perfect staffing standards.

4. Conclusion

So far, empirical studies on human resource management practices and corporate performance by domestic and foreign scholars have roughly shown two different conclusions: most scholars believe that human resource management practices have a positive correlation with corporate performance, but the conclusions of the research are not completely consistent. Most studies only prove that some human resource management practices have a significant correlation with corporate performance. This article divides the practice of human resource management into six modules functionally and separately studies the impact of the six modules on corporate performance. Another group of scholars believe that the impact of human resource management practices on corporate performance is not significant, or even believe that there is no positive or negative correlation between the two. The long-standing problems of overstaffing and inbreeding in state-owned commercial banks are overcome by establishing clear staffing standards and release employees who do not meet the standards, reducing the labor cost of the enterprise and improving the efficiency of talent utilization. The most important work of human resource planning is to establish and improve various personnel rules and regulations within the enterprise. Standardized systems can improve employee productivity and help enterprises to operate in an orderly manner. State-owned commercial banks need to better institutionalize the management and control of a large number of employees. The diversification of the intermediary mechanism shows that human resource management practices can affect corporate performance through a variety of factors and indirectly proves that there is a relationship between the two situations.

(1) As for which intermediary impact of the mechanism is more effective, and the company needs to consider it according to its own organizational structure and human resource characteristics. Different corporate characteristics, different corporate cultures, and even different regional cultures will affect the role of the intermediary mechanism. In summary, the research on the relationship between human resource management practices and corporate performance in the Chinese context is significantly different from Western empirical research conclusions. The empirical research conclusions of different domestic scholars are also very different, and the research objects are diverse. It is systematic and has little reference for the human resource management practices of domestic enterprises. It needs further research and discussion by scholars.

(2) Selection of training methods. The level of employee training of state-owned commercial banks determines the diversity of training methods. Staff training methods mainly include direct teaching, including lectures to teach theoretical knowledge, work guidance to employees, or job rotation to improve their work skills from practical operations. It can also take the form of brainstorming or outreach training to improve the comprehensive ability of employees. State-owned commercial banks should choose the appropriate method according to each training target and training purpose. For example, classroom-teaching methods should be adopted to introduce the basic situation of the company to new employees of the company, and the work rotation method and work guidance method should be adopted to allow employees to quickly master the skills required for their work. For high-level personnel or when a new problem needs solved, collective brainstorming can be used to brainstorm ideas.

(3) Training evaluation and risk prevention. Training evaluation is a necessary process to measure whether the training is effective after the training is over, to prevent the training from becoming a formality. They still regard the human resources department as a logistics support department of the enterprise organization and believe that the investment in human capital will only increase the labor of the enterprise. It is useful to make the right cost value for the enterprise. Even if individual corporate managers know that human resource management can have an impact on corporate performance, how human resource management affects corporate performance has become a "black box." In this context, whether the practice of human resource management will have an impact on corporate performance, how much impact it has and how this impact achieved, has become a common concern for theoretical and practical operators. Through the empirical research on the human resource management practice and corporate performance of state-owned commercial banks, the relationship between the two can further verified theoretically, and the domestic research on the relationship between the two can enriched. This article analyzes the problems of human resource management in our state-owned commercial banks and verifies the relationship between human resource management practices and corporate performance. The research shows that there is a significant positive correlation between the two. Therefore, to a certain extent, it can cause state-owned commercial banks to reflect on their own corporate human resource management practices and recognize.

Data Availability

The labeled data set used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was funded by the Youth Innovation Projects in Colleges and Universities in Guangdong Province, research on the impact of high-speed rail network on regional spatial optimization under the guidance of coordinated development, 2021WQNCX113.

References

[1] Z. Khan and S. Amin, "Bottleneck model with heterogeneous information," *Transportation Research Part B: Methodological*, vol. 112, no. 1, pp. 157–190, 2018.

[2] J. M. Cairney, K. Rajan, and D. Haley, "Mining information from atom probe data," *Ultramicroscopy*, vol. 159, no. 1, pp. 324–337, 2020.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9840753, 1 page
https://doi.org/10.1155/2023/9840753

Retraction

Retracted: Detection of DDoS Attack within Industrial IoT Devices Based on Clustering and Graph Structure Features

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Jing and J. Wang, "Detection of DDoS Attack within Industrial IoT Devices Based on Clustering and Graph Structure Features," *Security and Communication Networks*, vol. 2022, Article ID 1401683, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1401683, 9 pages
https://doi.org/10.1155/2022/1401683

Research Article

Detection of DDoS Attack within Industrial IoT Devices Based on Clustering and Graph Structure Features

Hengchang Jing and Jian Wang

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Correspondence should be addressed to Hengchang Jing; jhc@nuaa.edu.cn

Received 16 December 2021; Accepted 5 February 2022; Published 7 March 2022

Academic Editor: Chin-Ling Chen

Network available and accessible is of great importance to the Internet of things (IoT) devices. In this study, a novel machine learning method is presented to predict the occurrence of distributed denial-of-service (DDoS) attacks. Firstly, a structure of edges and vertices within graph theory is created to simultaneously extract traffic data characteristics. Eight characteristics of traffic data are selected as input variables. Secondly, the principal component analysis (PCA) model is adopted to extract DDoS and normal communication features further. Then, DDoSs are detected by fuzzy C-means (FCM) clustering with these features. In the case study, 2000 traffic data in dataset CICIDS-2017 are used to verify the practicability of this method. The results of recall, false positive, true positive, true negative, and false negative are 100.00%, 1.05%, 68.95%, 0.00%, and 30.00%. Compared with other methods, the results demonstrate that the detecting reliability is improved, and the method has a good effect on the detection of DDoS attacks.

1. Introduction

Network security problems have become increasingly outstanding with the development of the Internet of things (IoT) technology. There are a lot of malicious attacks on the network. Maintaining the stability and reliability of IoT devices is a complex task due to the highly distributed and multiple connected characteristics. Distributed denial-of-service (DDoS) attacks are the most common way to destroy the accessibility of a network. DDoS attacks have the characteristics of low launching cost and high attack intensity, which can cause significant harm to the victims quickly. The DDoS attack is different from a penetration attack, which does not invade the target servers by a Trojan or root program. DDoS attacks have two types, and one is a network protocol attack to damage servers by the network system vulnerability maliciously. The other is directly run out of resources by infinitely sending useless packages to the object [1, 2], which will lead the target system service to block, and the IoT equipment cannot provide a normal service or access to clients. The first type of attack can be effectively defended by system patching, but the second one must accurately distinguish legitimate traffic data from network flows. Thus, this dataset mining problem has drawn attention to many researchers in network security.

Service resources for the victims of DDoS attacks include network bandwidth, file system space capacity, open processes, or allowed connections [3]. These attacks will lead to the decrease in memory capacity resources, and bandwidth speed will inevitably decrease. According to the popularization of information technology, especially the IoT, more and more host types of botnets that is a host infected with a malicious program and under the control of an attacker appear [4]. Verizon revealed a DDoS attack on a US university, the campus network speed has slowed down significantly, and the domain name server (DNS) was flooded with abnormal queries from the school's approximately 5,000 IoT devices, including streetlights, vending machines, and other botnet devices [5].

It is usually hard for network security officers to identify them because many network devices such as routers, switches, and servers produce a vast amount of system log data. An effective way to track network status is to deploy monitoring agents in the network and collect log information corresponding to a change in system status [6]. Researchers have developed different models to address this problem, such as signature-based intrusion detection, entropy variation method, machine learning detection, and artificial intelligence-based method. Analyzing the correlation model is used to detect anomalous network activities through the temporal and process information [7]. A causal inference algorithm is developed to detect a nationwide research and education network in Japan by 15 months long system log messages collected [8]. A DDoS defense scheme for the IoT using dynamic population and point process theory is presented to predict and detect DDoS attacks by analyzing traffic data. A generalized entropy-based metric is proposed to detect the low rate DDoS attacks to the control layer [9]. RBF neural network is used as an anomaly based approach, and the detection ratio of 96% is shown in the UCLA dataset [10]. The clustering models, such as the K-means model [11] and Gaussian mixture density model [12], are unsupervised methods that classify datasets into multiple clusters only with varying distances of membership, which can divide each traffic data into different partitions for distincting DDoS and normal flows. The label will be not required in unsupervised methods. When the traffic data of network communication are divided into different partitions with clustering, the DDoS attacks will be easily found. However, the fuzzy C-means (FCM) cluster model in data mining is seldom used to perform DDoS detection.

For acquiring an effective detection method of DDoS attacks, this study proposes a novel detected method. The traffic dataset of network communication is first analyzed using graph theory. Then, the principal component analysis (PCA) is used to filter the characterization factors of DDoS attacks. The FCM clustering model divides the network flows of traffic data into different partitions. In the case study, the dataset of CICIDS-2017 was selected to verify the practicability of the method, and the results were presented. The novelty of this model is as follows: (1) the traffic data can be unsupervised for training, so labels are not needed; (2) using graph theory not only considers the topological structure relationship between IP and ports but also considers flows; and (3) many factors of traffic data can be automatically selected to reduce the overload of calculation and improve the accuracy of clustering.

2. Graph Structure Features

Graph theory [13–15] is used to build a topological structure of traffic data. The traffic data can be abstracted as a directed graph (DG) in communication networks. The communication relationship, frequency, flow duration, and other valuable information between vertices could be regarded as the edges (links) $E = \{e1, e2, \ldots, em\}$ and the IP addresses and ports are vertices (nodes) $V = \{v1, v2, \ldots, vn\}$, where m is the total number of edges and n is the total number of vertices.

For instance, six vertices exist in a graph structure containing $v1$ = 172.16.0.1 : 43201, $v2$ = 192.168.10.50 : 80, $v3$ = 101.69.185.208 : 443, $v4$ = 192.168.10.16 : 51784, $v5$ = 103.43.91.16 : 443, and $v6$ = 192.168.10.9 : 9901, to which three edges connect ($e1$ = 172.16.0.1 : 43201 ⟶ 192.168.10.50 : 80, $e2$ = 101.69.185.208 : 443 ⟶ 192.168.10.16 : 51784, and $e3$ = 103.43.91.16 : 443 ⟶ 192.168.10.9 : 9901).

The weight of the edge contains various information that can be expressed as an array. The connectivity of traffic data can be considered an adjacent matrix A to show the relationship between these IP addresses and ports clearly. The matrix A is as follows:

$$A = \begin{bmatrix} \mathbf{w}_{11} & \cdots & \mathbf{w}_{1j} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{i1} & \cdots & \mathbf{w}_{ij} \end{bmatrix}, \tag{1}$$

where the vector wij represented the array of weights between nodes i and j. If the nodes i and j are connected, the weights are nonzero. Otherwise, it is zero.

The weights are the traffic data features. Different features can reveal various communication relationship characteristics in the topological structure. The DDoS attack contains directed attack and reflected attack [16, 17]. In the reflected DDoS attack, attackers indirectly attack the target IP service and send specialized packet data to an opening server for disguising IP address, and the opening server will reply to the request packet data sent to the attacked server many times. It is difficult to judge a DDoS attack only by its IP address and ports. However, a DDoS attack is from one source address to a terminal address to break down servers and have diverse characteristics. Thus, the features of edges can recognize attacks effectively. The DDoS attack should also be distinguished from the flash crowd that is a normal access behavior of the clients. Flash crowd appears when a huge number of clients access a server simultaneously due to top search results, popular products, and so on. Users want to get interested in information from the server as soon as possible. The server is slow or even shut down, which is unexpected, and most do not want to see it in advance. Overall consideration, we analyzed graph-based and flow-based features under the DDoS attack environment to select features to detect DDoS attacks. Eight features are selected, as follows.

2.1. Total Forward Packet. The forward packet means a request sends from a source node to the target node. In traffic data, the total forward packet represents the number of received data packets of the target node from an adjacent source node in the network. The total forward packet can be regarded as an indicator of the activity of a source node. Useless information and command send to slaves from masters in the DDoS attacks.

2.2. Total Backward Packet. The backward packet is the reply information sent to the source node after the target node receives a request. The total backward packet represents the number of data packets sent from a target node to an adjacent source node in the network. In detecting DDoS attacks, it can represent the slaves' activity of the network.

2.3. Standard Deviation of Backward Packet Length. The standard deviation of backward packet length represents the fluctuation of packets replying from a target node to a source node. The standard deviation of the backward packet length of the DDoS attack is smaller than normal traffic. In DDoS attacks, the length of packets between two particular nodes is all the same, and the interval time tends to be stable. The standard deviation is almost zero or the same small size. Thus, the length of packets is the same when the message of the victim sever returns to the attack node. In normal traffic, the length of packets fluctuates significantly due to different requests. Thus, the standard deviations are large and variable to the different connection nodes.

2.4. Total Visit View. The total visit view is the number of accesses to a destination IP and port from a source node continuously. In the DDoS attacks, the source node will continue sending packets to disrupt normal traffic on the target server until managers detect it.

2.5. Average Packet Length. The average packet length is a statistical value of a packet in a duration of time. In DDoS attacks, the average packet length is small because the duplicate packets only contain header files without any data fields or less content. Each data packet has the same header but different contents in normal flows. The average length of the packet is large and various.

2.6. Flow Duration. The flow duration is the total communication time between two nodes from connection to disconnection. The flow duration of packets sent by the same attacker tends to be stable in the DDoS attacks, while the duration time frequently fluctuates in normal communication.

2.7. Standard Deviation of Flow Interval Time. The flow interval represents the interval between sending each packet during a flow. When the DDoS attackers send packets, the interval time of flow tends to be equal. However, the interval time of normal flows depends on the reply time of the target server. The destination vertices receive different packets, and the processing time is also different. Furthermore, the interval time of normal traffic is affected by noise, network bandwidth, receiving window size, sending window size, etc., which shows a significant difference from DDoS attacks [5]. In addition, the interval time of normal traffic is limited by network bandwidth, noise, size of sending window, and other factors, which is significantly different from DDoS attacks.

2.8. Mean Active Time of Flow. The meaning of active mean is different from the traffic duration mentioned above. It represents the survival time of each packet sent within the communication time of two vertices. The definition is the total interval between sending the connection request packet and the last disconnect request packet. DDoS attackers make attacks many times in a short period, and the sending packet is generally the same and small. On the contrary, the normal flow survival time depends on the communication time. Otherwise, the value is zero.

The label of each traffic data is normal communication or DDoS attack. For detecting convenience, the labels are digitized (zero for normal communication and one for DDoS attacks). In the dataset, each flow has been labeled based on its weights.

For the example above, *e1* is assumed as a DDoS attack. *e2* and *e3* are normal communication.

Then, the nodes *v1-v6* are connected by *e1-e3*, and the weight array is as follows:

$$A = \begin{bmatrix} 0 & \mathbf{w}_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{w}_{34} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{w}_{56} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad (2)$$

$$\mathbf{w_{ij}} = \left[w_{ij}^1, w_{ij}^2, w_{ij}^3, w_{ij}^4, w_{ij}^5, w_{ij}^6, \ldots, w_{ij}^k\right]^T,$$

where the w_{ij}^k is the kth weight between nodes i and j.

The eight features ($k = 8$) can be inputted into the weight array. The values of features are assumed as known. Then, the $\mathbf{w_{ij}}$ can be written in Table 1, where the k is in keeping with the above orders.

The label array can be expressed as follows:

$$Y = \left[y_1, y_2, y_3\right]^T, \qquad (3)$$

where Y is the array of label between nodes i and j; y_i is the label of ith edge.

With respect to the assumption, the label array can be written as $[1, 0, 0]^T$.

3. Dimensionality Reduction in the Weight Matrix

The PCA is commonly applied for dimensionality reduction, which projects data onto only the first few principal components to obtain lower-dimensional data [18]. Thus, PCA can be solved by lossy compression of a dataset to express characteristics by less dimensional data.

A group of new orthogonal bases should be found in the PCA algorithm where the projection's data have a maximum variance value. In other words, the distance of data is the largest in the projection of orthogonal basis. When the m weights: $\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w_m}\}$ exist, and each weight has n dimensions: $\mathbf{w_i} = [w_i^1, w_i^2 \ldots, w_i^n]^T$, the variance of all data projected onto that basis can be expressed as [19]follows:

$$J_j = \frac{1}{m}\sum_{i=1}^{m}\left(\mathbf{w_i^T u_j} - \overline{w}_\mathbf{i}\mathbf{u_j}\right)^2, \qquad (4)$$

Table 1: Weight array of three edges.

Edge	w^1	w^2	w^3	w^4	w^5	w^6	w^7	w^8
e1	3	4	5795.50	7	1661.86	77116	30796.08	1000
e2	2	3	0	1	13.60	655938	327914.33	0
e3	2	0	0	1	9	377	0	0

where m is the number of weight samples; $\mathbf{w_i}$ is the ith weight after the zero-mean initialization; $\overline{w}_\mathbf{i}$ is the average weight; $\mathbf{u_j}$ is the jth orthogonal basis; and J_j is the variance when the dataset projects onto the orthogonal basis j.

Then, the zero mean is processed for each element of X by (5). The X columns are centered on having an average value zero and scaled to have a standard deviation one.

$$\mathbf{w_i} = \frac{\mathbf{w_i} - \overline{w}_\mathbf{i}}{\sigma}, \tag{5}$$

where σ is the standard deviation of weight array $\mathbf{w_i}$.

$\overline{w}_\mathbf{i}$ is zero when zero-mean initialization is processed. Then, (4) can be written as [19]follows:

$$\begin{aligned} J_j &= \frac{1}{m}\sum_{i=1}^{m}\left(\mathbf{w_i^T u_j}\right)^2 \\ &= \mathbf{u_j^T}\frac{1}{m}\sum_{i=1}^{m}\left(\mathbf{w_i w_i^T}\right)^2 \mathbf{u_j}. \end{aligned} \tag{6}$$

The matrix form can be expressed as follows:

$$\begin{aligned} J_j &= \frac{1}{m}\mathbf{u_j^T}\mathrm{XX}^T\mathbf{u_j} \\ &= \frac{1}{m}\mathbf{u_j^T} S \mathbf{u_j}, \end{aligned} \tag{7}$$

where X is the matrix of weights, and the equation is shown in (8); S is the value of matrix multiplication between X and X^T, which is also called the covariance matrix.

$$\begin{aligned} X &= \begin{bmatrix} \mathbf{w}_1 \\ \cdots \\ \mathbf{w_m} \end{bmatrix} \\ &= \begin{bmatrix} w_1^1 & \cdots & w_1^n \\ \vdots & \ddots & \vdots \\ w_m^1 & \cdots & w_m^n \end{bmatrix}. \end{aligned} \tag{8}$$

The orthogonal basis can be deviated by the Lagrangian operator [20]. For obtaining an orthogonal basis, the maximum variance of the data projected onto the basis is equal to the eigenvalue of the covariance matrix of X. It can be written as follows:

$$\max J_j = \lambda_j \tag{9}$$

When the dimensionality reduction is processed, the eigenvalues are first arranged in descending order. The weight matrix of reduced dimension can be solved by the eigenvectors corresponding to the first k maximum eigenvalues of the covariance matrix if the dimension reduces to k. In order words, the orthogonal basis is equal to the eigenvectors of the covariance matrix of X.

With respect to the definition of covariance, the covariance of matrix X can be expressed as [21]follows:

$$\begin{aligned} \mathrm{cov}(X) &= \frac{1}{m}\mathrm{XX}^T \\ &= S, \end{aligned} \tag{10}$$

where cov represents the covariance matrix.

Then, the covariance matrix S is diagonalized, and the eigenvectors and eigenvalues can be obtained. Thus, the matrix of dimensionality reduction can be calculated by (9).

$$X_{\mathrm{new}} = \begin{bmatrix} \mathbf{u}_1 \\ \cdots \\ \mathbf{u_k} \end{bmatrix}^T X, \tag{11}$$

where X_{new} is the matrix of $k * m$.

The above example is further used to instruct, and the first weight arrays are selected for simplifying the weight matrix X to express the process clearly. It can be written as follows:

$$\begin{aligned} X &= \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 4 \\ 2 & 3 \\ 2 & 0 \end{bmatrix}. \end{aligned} \tag{12}$$

After the zero-mean initialization, the weights of three edges values are shown in Table 2.

Covariance matrix S is solved by $1/m\mathrm{XX}^T$, shown in Table 3. Then, the eigenvectors and eigenvalues of S can be obtained, shown in Table 4.

The weight matrix should be reduced from n dimensions to k dimensions. Thus, an appropriate k value ought to be determined. The general selection criterion is the proportion of variance before and after projection. The higher proportion will have a higher correlation, so they are used as the selection criterion of the k value.

With respect to the relationship between covariance and eigenvalue, it can be expressed as [22]follows:

$$\begin{aligned} q &= \frac{J(X_{\mathrm{new}})}{J(X)} \\ &= \frac{\sum_{j=1}^{j}\lambda_j}{\sum_{j=1}^{n}\lambda_j}, \end{aligned} \tag{13}$$

where q is the expectation value.

However, not just only one array is selected. A higher proportion q will have a higher correlation. Thus, a series of arrays with large expectation values are used. The sum of the expectation value q is larger than 90% with respect to the analysis of some references [5], [23-25].

Table 2: Zero-mean initialization weights of three edges.

Edge	w^1	w^2
e1	0.67	1.67
e2	-0.33	0.67
e3	-0.33	-2.33

Table 3: Covariance matrix of three edges.

	1	2
1	0.22	0.56
2	0.56	2.89

Table 4: Eigenvalue and eigenvector of three edges.

Eigenvector	1	2
1	0.22	0.56
2	0.56	2.89
Eigenvalue	0.11	3.00

In this example, the expectation values are 96.42% and 3.57%, respectively. Thus, the second weight array can be ignored. The dimension can be reduced to one. Finally, a new weight matrix (3 $*$ 1) can be solved by (9). It can be written as follows: $[1.08, 0.30, -1.38]^{\mathrm{T}}$.

4. Fuzzy C-Means (FCM) Clustering

After the PCA dimensionality reduction, cluster analysis can be processed with respect to the new weight matrix. A fuzzy C-means (FCM) is an unsupervised learning model presented in 1973 [26, 27], which does not require manual creation of categories for dataset labels. The FCM algorithm is an effective cluster model based on a fuzzy clustering algorithm to minimize an objective function, dividing data into different classes by the degree of membership. It is widely applied in different areas, such as news classifying, user buying patterns (cross-selling), image segmentation, and genetic technology. However, it is seldom used to classify nodes to normal access and DDoS attack in the network security area. Therefore, this study applies the FCM to judge the DDoS attack.

The weight analysis matrix (11) is used as a sample observation matrix to divide each edge into different partitions. The number of partitions c is determined manually, and a membership matrix $\mathbf{M}$ is generated randomly, where the number of matrix rows is the same as the number of partitions (total of c classes) and the columns are equal to the index of edges, a total of m (such as three in the example of above). When the number of dimensions of sensitivity is assumed as n, the membership matrix $\mathbf{M}$ can be expressed as follows [28]:

$$\mathbf{M} = \begin{bmatrix} \Delta\mathbf{M}_1 \\ \vdots \\ \Delta\mathbf{M}_m \end{bmatrix} = \begin{bmatrix} M_{11} & \cdots & M_{1c} \\ \vdots & \ddots & \vdots \\ M_{m1} & \cdots & M_{mc} \end{bmatrix}^T, \tag{14}$$

where M_{ij} is the membership of edge i at the partition j, and the membership values in the membership matrix are all ranged from 0 to 1. The membership represents the degree of reliability of an edge in a partition.

Then, the center of partitions $\mathbf{C}_j$ $[\mathbf{C}_j = (C_{1j}, C_{2j}, \ldots, C_{Mj})]$ in each class is determined as follows:

$$\mathbf{C}_j = \frac{\sum_{i=1}^{N} M_{ij}^m \boldsymbol{\omega}_j}{\sum_{i=1}^{N} M_{ij}^m}, \tag{15}$$

where m is a power exponent m $(m > 1)$.

With respect to the center of clustering, the membership matrix can be revised via solving the Euler distance [27]:

$$\mathbf{M}_{\mathbf{ij}} = \frac{1}{\sum_{k=1}^{c} \left(d_{ij}/d_{ik} \right) 2/m - 1}, \tag{16}$$

where d_{ij} is the Euler distance of edge i at the partition j; represents the distance solving equation that can be expressed as follows:

$$\boldsymbol{\omega}_j - \mathbf{C}_j = \sqrt{\sum_{i=1}^{N} \left(\boldsymbol{\omega}_{\mathbf{j}} - \mathbf{C}_{ij} \right)^2}. \tag{18}$$

Then, an objective function is employed to solve the weights that are the sum of squares for the distance sensitivity values to their cluster centers, expressed in (19). The objective function should be minimized and the partition of the minimum value isselected as their divided clusters [29].

$$F(\mathbf{M}, \mathbf{C}) = \sum_{i=1}^{N} \sum_{j=1}^{c} M_{ij} d_{ij}^2, \tag{19}$$

where F is the objective function that should be optimized; C is the matrix of the center of partition that can be expressed as $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_c]^T$.

It is not easy to decrease the convergence value to zero in the numerical calculation. Thus, a convergence condition ξ $(\xi > 0)$ can be set to judge to stop the looping. Meanwhile, a maximum iteration time is also set to prevent an endless loop. The convergence condition can be expressed as follows: $|F^l(\mathbf{M}, \mathbf{C}) - F^{l-1}(\mathbf{M}, \mathbf{C})| < \xi$, where l is the lth iteration time. (15)–(19) are repeated until the result is up to the convergence condition or maximum iteration time minimizes the objective function. Finally, the objective function up to the minimum and the final membership matrix is obtained. The edges of IP and port connections are all classified.

The above example can be classified into different partitions using the FCM algorithm. When c assumes two, the three edges will be divided into two partitions. The first partition only contains one edge, *e1*. The other contains two: *e2* and *e3*. The label of *e1* is DDoS. *e2* and *e3* are normal communications. Thus, a similar dataset can be separated into different partitions.

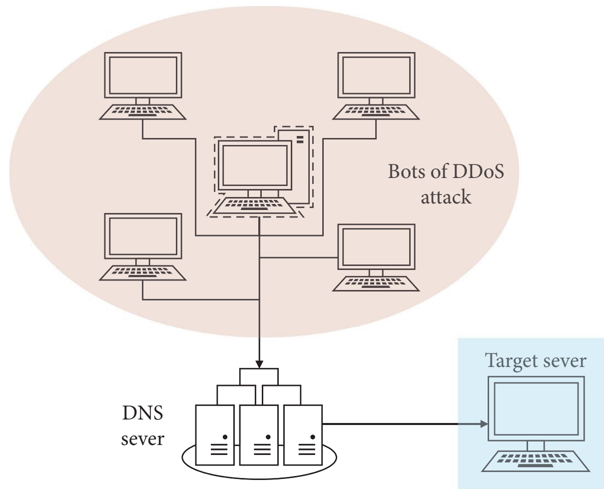


Figure 1: DDoS attack process.

5. Case Study

CICIDS-2017 dataset is employed to verify the practicability of this method [30, 31]. The dataset contains benign and the most up-to-date common attacks, which resembles the actual real-world data (PCAPs). Cases of high frequently used network flows are employed, including the traffic data of benign and DDoS attacks. It also includes the results of the network traffic analysis using CICFlowMeter with labeled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols, and attack [31]. The DDoS attacks were implemented on Friday afternoon and captured in the dataset, which has a total of 225,747 flows, of which more than 40,000 DDoS attack flows. According to the official label, there are three bots, and their IP addresses are 205.174.165.69, 205.174.165.70, and 205.174.165.71. The network firewall IP addresses are 205.174.165.80 and 172.16.0.1, and the victim host IP address is 192.168.10.50. The flowchart of DDoS attacks process is shown in Figure 1.

These traffic data are set to a CSV file. Every flow has 83 properties in the CSV files, such as the timestamp, source, and destination IPs, source and destination ports, and flow duration, while one label exists, which can represent the DDoS attack or normal communication. Besides, the computer devices are as follows: CPU is i7-9700K; RAM is DDR4-96G; ROM is Intel SSD 1T; the operating system is Windows 10; and MATLAB 2018b is used.

Then, the combined PCA and FCM algorithm is used to create a DDoS detection model. The process of clustering is illustrated in Figure 2.

At first, the direct graph (DG) model is created within graph theory to reveal traffic data characteristics for both the victims and bots to generate the relationship between source and destination IP port structure. In the DG model, the vertices are presented by the combination of IP and ports. Two vertices directly point to the destination IP port from the source, called the edges. Some properties of edges can be selected as the input variables for detecting DDoS attacks. The most obvious characteristics are total forward packet, total backward packet, the standard deviation of backward packet length, total visit view, average packet length, flow duration, the standard deviation of flow interval time, and the mean active time of flow, which are extracted as the input variables to generate a matrix A by (1).

However, only edges can be found in the CSV file. The preprocessing is measured to extract the information of vertices by MATLAB. The second and third columns are the source IP and port, respectively. The content of two columns is extracted and combined as a node of DG. Meanwhile, the fourth and fifth columns are the destination IP and port. These two columns are also combined as a node pointed by the source IP port. In this study, a total of 2037 nodes are selected as graph-based features. These nodes can formulate 2000 edges, including 600 DDoS attack edges and 1400 normal edges (3 : 7).

Then, the eight characteristics are converted into edge features in a weight matrix X. The matrix X of dimensionality reduction in eight features should be solved in the PCA processing. To further reduce X's dimension, an appropriate k value should be determined. The proportion of the selection criterion of k value is solved by (9), and the results are shown in Table 5.

When the weights are three, the total proportion is 98.48%. These preceding three weights can be regarded as the essential factors for predicting DDoS attacks. Thus, only total forward packet, total backward packet, and standard deviation of backward packet length are retained. A new three-dimensional weight matrix can be solved by (11).

At last, the FCM clustering algorithm is employed to predict the DDoS flows. In this study, FCM clustering was performed within different partition values c. The power

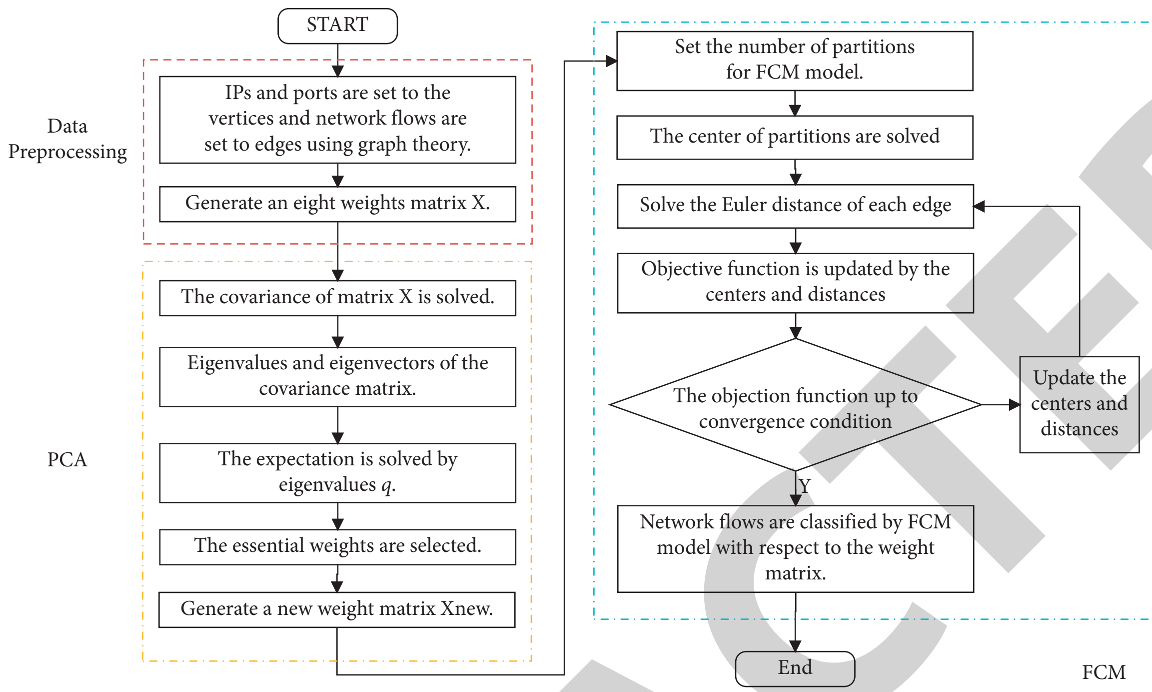


Figure 2: Flowchart of the clustering.

Table 5: Expectation values of the weight matrix.

Weight	w^1	w^2	w^3	w^4	w^5	w^6	w^7	w^8
Proportion (%)	93.83	3.13	1.51	0.76	0.58	0.14	0.05	0.00
Total proportion (%)	100.00	6.17	3.04	1.52	0.77	0.19	0.05	0.00

Table 6: Number of detection of DDoS attacks in the PCA-FCM.

Partition (c)	Normal communication	DDoS attack
2	1202	798
3	1377	623
4	1379	621
5	1367	633

Table 7: Detection efficiency with different c values in the PCA-FCM.

Partition (c)	Number of attack node	Recall (%)	False-positive rates (%)	True-positive rates (%)	True-negative rates (%)	False-negative rates (%)
2	600	100	9.90	60.10	0.00	30.00
3	600	100	1.05	68.95	0.00	30.00
4	600	100	1.75	68.25	0.00	30.00
5	600	100	1.65	68.35	0.00	30.00

index is set to 2.0, the tolerance is 10^{-5}, and the maximum iteration time is 100. In the clustering of different c values tested, all c values detect the attack edges, and the lowest false alarm rate is when $c = 3$. The center of three partitions is (0.45, 0.01, 0.01), (0.08, 0.02, -0.01), and (0.78, 0.02, -0.01). The results of detection efficiency are shown in Table 6, and

Table 8: Predicted results for the CICIDS-2017 dataset.

Method	Partition (c)	Dimension	Recall (%)	False positive (%)	True positive (%)	True negative (%)	False negative (%)
FCM	3	8	100	1.20	68.80	0.00	30.00
PCA-FCM	3	4	100	1.05	68.95	0.00	30.00
K-means	2	8	100	10.25	59.75	0.00	30.00
K-means	3	8	100	1.20	68.80	0.00	30.00
K-means	4	8	100	1.65	68.35	0.00	30.00
K-means	5	8	100	1.20	68.80	0.00	30.00
PCA-K-means	2	3	100	10.35	59.65	0.00	30.00
PCA-K-means	3	3	100	10.20	59.80	0.00	30.00
PCA-K-means	4	3	100	1.65	68.35	0.00	30.00
PCA-K-means	5	3	100	1.70	68.30	0.00	30.00
NMF-K-means [5]	2	4	100	4.80	65.20	0.00	30.00
NMF-K-means [5]	3	4	100	6.20	63.80	0.00	30.00
NMF-K-means [5]	4	4	100	2.03	67.97	0.00	30.00
NMF-K-means [5]	5	4	100	2.34	67.66	0.00	30.00

the recall, true-positive, true-negative, false-negative, and false-positive rates are shown in Table 7. No one DDoS attack is missed.

6. Comparisons

For verifying the effectiveness of this method for clustering the traffic data to normal and DDoS flows, the K-means clustering algorithm, and nonnegative matrix factorization (NMF), dimensional reduction model can be employed in the above case [32]. NMF was proposed in 1999, which makes all components after decomposition nonnegative, and at the same time realizes nonlinear dimension reduction. It corresponds to the intuitive understanding that the whole is made up of the parts, so it captures in a sense the nature of intelligent data description. Meanwhile, the pure clustering methods that the dimension is not reduced are also used to compare. The results of the two methods after optimization are shown in Table 8.

The PCA-FCM model in this study greatly affects DDoS attacks in network communications, compared with other methods. The recall rate, true negative rate, and false negative rate are 100%, 0%, and 30%, respectively, indicating that all DDoS attacks are detected. The false positives have decreased to 1.05%, while true positives have increased to 68.85%, when partition and dimension are 4 and 3, respectively.

7. Conclusion

This study presents a novel PCA-FCM model to detect DDoS attacks where the topological structure is taken into account between IP ports of source and destination. Then, characteristics, including total forward packet, total backward packet, the standard deviation of backward packet length, total visit view, average packet length, flow duration, the standard deviation of flow interval time, and mean active time of flow, are considered input variables for clustering. The PCA model is employed to reduce the dimensions of features further. Then, the bots are detected by FCM clustering with these features. The CICIDS-2017 dataset is employed to verify this method in the case study. The results demonstrate that the method has a high detecting reliability. The PCA-FCM method is suitable for DDoS detection. The recall, true negative, and false negative are 100.00%, 0.00%, and 30.00% that means no one DDoS attack is missed. The false positive and true positive are 1.05% and 68.95% compared with FCM, which has a considerable improvement.

With respect to the results, the PCA-FCM model has three advantages. Firstly, FCM uses unsupervised training and does not require labels; secondly, the topological structure relationship between IP and ports is connected by a DG structure. Thirdly, input variables can be automatically selected by PCA within many factors of dataset to reduce the overload of calculation. Therefore, this method provides a new horizon to network security.

However, some disadvantages can be discovered. Firstly, the vertex property is not considered an input variable for clustering. Secondly, the number of partitions should be calculated automatically in the clustering algorithm. Thirdly, a supervised model can be applied further to recognize new data after clustering by the edges label. Therefore, further research and improvement of this method should be conducted in the future to accurately and quickly detect DDoS attacks.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] A. Bhardwaj, G. V. B. Subrahmanyam, V. Avasthi, H. Sastry, and S. Goundar, “DDoS attacks, new DDoS taxonomy and mitigation solutions - a survey,” in *Proceedings of the 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pp. 793–798, Paralakhemundi, India, October 2016.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9895359, 1 page
https://doi.org/10.1155/2023/9895359

Retraction

Retracted: Digital Protection Technology of Cultural Heritage Based on ArcGIS Geographic Information Technology Algorithm

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Guo, W. Jiang, Q. Zhang, and K. Wang, "Digital Protection Technology of Cultural Heritage Based on ArcGIS Geographic Information Technology Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 3844626, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3844626, 10 pages
https://doi.org/10.1155/2022/3844626

Research Article

Digital Protection Technology of Cultural Heritage Based on ArcGIS Geographic Information Technology Algorithm

Xiaoying Guo [1], Weigang Jiang [2], Qian Zhang [1], and Kai Wang [3]

[1] *Department of Animation, School of Art and Design, Zhengzhou University of Light Industry, Zhengzhou 450003, China*
[2] *Henan Vocational College of Art, Zhengzhou 450011, China*
[3] *Zhengzhou University of Technology, Zhengzhou 450003, China*

Correspondence should be addressed to Xiaoying Guo; 2003033@zzuli.edu.cn

Received 29 December 2021; Revised 24 January 2022; Accepted 28 January 2022; Published 7 March 2022

Academic Editor: Chin-Ling Chen

In order to improve the effect of digital protection of cultural heritage, this paper analyzes the spatial and temporal characteristics of cultural heritage with the support of ArcGIS technology, conducts research on ArcGIS' geographic information technology algorithm, and improves the algorithm with the support of digital technology. Moreover, this paper combines the watermark algorithm to digitally mark cultural heritage to improve the copyright effect of cultural heritage. In addition, this paper proposes a topological integrity verification method based on weak watermark vector geographic data to improve the reliability of digital processing of cultural heritage, and build an intelligent cultural heritage digital protection system model. The research results show that digital protection technology of cultural heritage based on the ArcGIS geographic information technology algorithm in this paper can play an important role in the digital protection of cultural heritage.

1. Introduction

With the rapid progress and renewal of social productivity and production methods, global economic models, values, and thinking models are all close to convergence, and the traditional culture, beliefs, living habits, and concepts belonging to their respective countries and nations are gradually being ignored. Looking back at the domestic situation as well, the past few decades have been an era of rapid economic development. However, in the process of economic development, the protection of and attention to the country and national culture have been neglected, and many precious national treasures, cultural and traditional skills, etc., have been destroyed or even disappeared. Due to the continuous expansion and development of urban land, the traditional rural culture based on farming civilization has been swallowed or shrunk. The mismatch between traditional cultural heritage and modern life and functions has caused a crisis of protection and inheritance. As a result, the reflection of opposing cultural convergence and protecting regional culture has become a new social demand of domestic and foreign governments and scholars, and has stimulated the exploration and maintenance of their own national and regional cultural activities.

Intangible cultural heritage tourism has become an important way for tourists to understand Chinese history, culture, and folk customs. At the same time, the impact of intangible cultural heritage tourism on heritage sites has also attracted widespread attention from all sectors of society. As the contradiction between intangible cultural heritage protection and tourism development continues to sharpen, the demands of relevant stakeholders cannot be met, and the quality of tourist experience is affected. In the process of its integration with tourism, problems such as single type of tourism experience, similar content of tourism experience, and low participation in tourism experience have emerged. Therefore, it is necessary to carry out further empirical research on tourist experience, using empirical research methods and concepts, to explore the problems existing in the intangible cultural heritage tourism experience.

As an intangible spiritual and cultural resource, intangible cultural heritage is the root that nourishes the soul of the Chinese nation. At present, intangible cultural heritage tourism is still at the stage of knowledge reserve. The maturity of theoretical research can provide guidelines for the future development of intangible cultural heritage tourism.

According to the digital protection requirements of modern cultural heritage protection, this paper studies the geographic information technology algorithm of ArcGIS, and improves the algorithm with the support of digital technology to provide technical support for the digital protection of cultural heritage.

The organizational structure of this paper is as follows: the introduction of the first part mainly describes the necessity and status quo of the digital protection of cultural heritage; the second part summarizes and analyzes the academic research status of the digital protection of cultural heritage through literature review; the third part is about the cultural heritage. The research on the algorithm of the key technology of digital protection provides the algorithm basis for the construction of the subsequent intelligent model. The fourth part builds the digital protection model of cultural heritage, and verifies the effect of the model proposed in this paper. The effective results are summarized.

2. Related Work

Regarding the research on the concept of intangible heritage, foreign scholars have experienced the process from protecting cultural property to attaching importance to the common heritage of mankind, from maintaining material cultural heritage to understanding the importance of intangible cultural heritage as the heritage of human civilization, and their research results are rich [1]. Literature [2] comprehensively defines the theoretical conceptual system of cultural heritage by comparing and analyzing the conceptual connotation of cultural heritage and cultural property, exploring the content of cultural awareness and cultural rights, etc.; Andrioti et al. [3] clarify that intangible cultural heritage is different from the key characteristics of tangible cultural heritage and the internal connection between the two also point out the problems and countermeasures in the protection of intangible cultural heritage. Lin [4] recounted the evolution of the concept of intangible heritage, and studied the standards of intangible heritage protection and related factors.

Intangible cultural heritage is a unique achievement of human civilization with multiple values. Current academic research on the evaluation of intangible cultural heritage has paid more attention to its social and economic aspects. Runhao [5] uses the contingent valuation method and the cost-benefit method to propose measures for the revival of cultural heritage. Yakar and Doğan [6] proposed that by displaying the value of intangible cultural heritage, it played a role in coordinating the conflict between North Korea and South Korea. Aparac-Jelušić [7] proposes the mutual relationship between intangible heritage values, which promotes urban development and the survival of cultural resources by expanding the tourism value of heritage, and enhances its social value. Khan et al. [8] proposed a cultural heritage protection model based on heritage values.

Research on the establishment of a policy system for protection management. Rahaman and Kiang [9] clarify the importance and urgency of intangible heritage protection, and propose that intangible heritage protection management should formulate detailed protection planning schemes, formulate and implement protection and development rules and regulations, and emphasize the scientific and orderly protection of intangible heritage. Mohammed Mahmoud Mohammed Ahmed [10] puts forward a series of protection measures based on the analysis of the status quo of intangible heritage, the exploration of the problems and causes, and realizes the sustainable development of the inheritance of intangible heritage protection. Regarding the development and utilization of intangible cultural heritage, the academic circles mainly focus on the model, development benefits, and intangible cultural heritage protection and utilization plans for individual cases. Kelly [11] pointed out that intangible cultural heritage is an important resource that stimulates social and cultural innovation and should be used to give full play to its value. Ulvi [12] believes that due to the complex social environment and the disappearance of traditional boundaries, the preservation and tourism status of intangible cultural heritage is inseparable from the local environmental factors, and the protective development of intangible cultural heritage should be based on intangible cultural heritage tourism in the local environment. Features are recognized. Rahaman [13] takes literary tourism cultural heritage as the research direction and analyzes that most of the literary tourism cultural heritage originates from the people's loyalty to literary sacred sites. Both sense of sight. Sciacchitano [14] believes that an attractive and market-compliant development model should be established to promote the planning of cultural heritage tourism.

Tang et al. [15] expounds the origin of the concept of intangible heritage, the understanding of the definition and category of intangible heritage, and puts forward its own views on the concept of intangible cultural heritage field, the key protection objects of intangible cultural heritage, and the issue of "cultural rights." Vučković et al. [16] analyze the two major types of intangible cultural heritage, namely, morphological culture and behavioral culture, of which morphological culture is the core part of intangible cultural heritage. Bec et al. [17] interpret and reconstruct the concept of intangible cultural heritage, expounds the connotation and value of the concept, and emphasizes the strengthening of research on living literary heritage. Deniz et al. [18] pointed out that the characteristics of intangible cultural heritage, such as the personality, cultural heritage, and nationality, mean its unique value. Champion and Rahaman [19] analyze the characteristics of the economic value of intangible cultural heritage and believes that attaching importance to the economic value of intangible cultural heritage is of great significance to the realization of national characteristics and the creation of new economic growth points.

3. Key Technologies for Digital Protection of Cultural Heritage

In this paper, the digital watermarking algorithm is used in the digital protection of cultural heritage, and the research on the digital watermarking algorithm is carried out first.

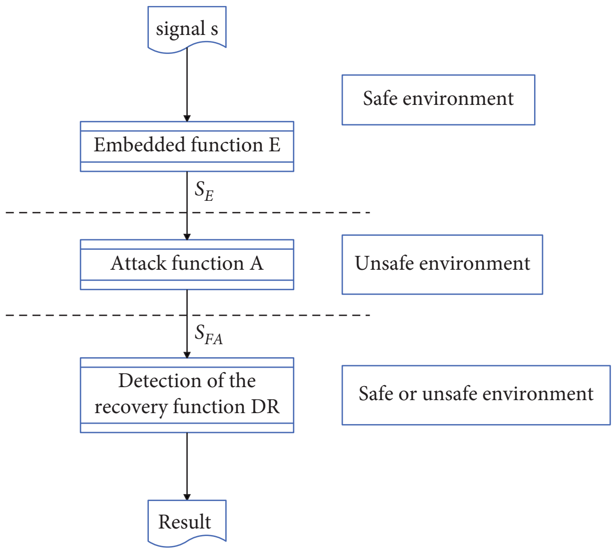


FIGURE 1: The life cycle of a digital watermark.

Digital watermarking has been an important research direction of information hiding technology, and it is an effective way to realize source identification, identity verification, and copyright protection. Digital watermarking technology refers to embedding secret watermarking information in digitized data content. This technology closely combines the watermark with the source data and hides it to make it an inseparable part of the source data. From this, it determines the copyright owner, verifies the ownership, tracks infringements, verifies the authenticity of the source of the digital content, identifies the purchaser, provides other additional information about the digital content, and so on.

Digital watermarking technology is a cutting-edge technology developed in the field of information security. Its characteristic is that it does not affect the use value of the original carrier, nor is it easy to be detected and modified again, but it can be identified and identified by the manufacturer.

Generally, the life cycle of a digital watermark consists of three parts: the watermark embedding phase, the data attack phase, and the watermark detection and recovery phase. Figure 1 shows the life cycle of digital watermarking. The information embedded in a signal is called a digital watermark, and the signal embedded in a watermark is called the host signal. A watermarking system is usually divided into 3 independent steps: embedding, attacking, and detecting. In the embedding process, an algorithm is needed that can process the carrier signal and watermark information at the same time and can generate a watermarked signal. Then, the algorithm can propagate or store the watermarked digital signal. If the data are modified during the propagation process, it is called an attack. However, the modification may not be malicious. The term "attack" is derived from copyright protection applications, because a third party may modify the digital watermark to remove it. Moreover, there are many ways to modify, such as data compression, image or audio cropping, or random increase of noise. The detection (or extraction) algorithm attempts to extract the watermark from the attacked signal. If the signal is not modified during the propagation, the watermark still exists and can be extracted. In robust watermarking, even if the modification range is large, the watermark information should still be accurately extracted. In fragile watermarking, if the signal is modified in any way, the watermark information cannot be extracted.

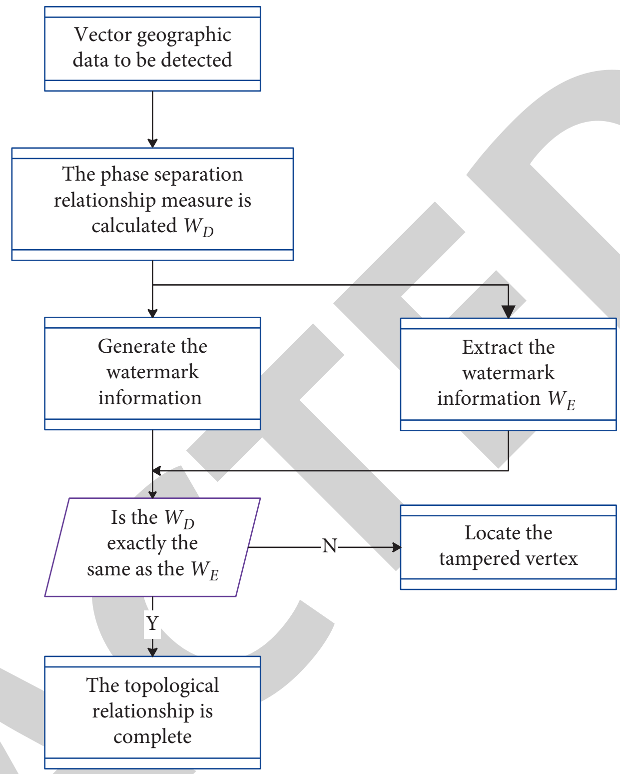


FIGURE 2: Checking process of topology integrity.

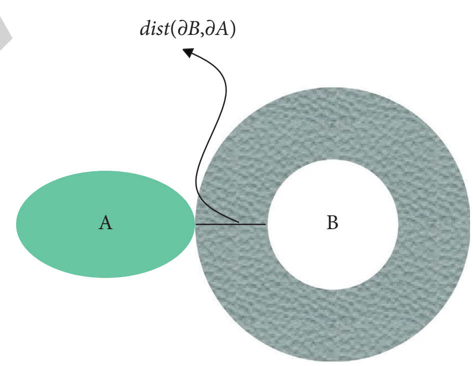


FIGURE 3: Schematic diagram of the calculation of the detachment relationship measurement value.

Digital watermarking has different classification methods according to its different application purposes. According to the antiattack of digital watermarking, it can be divided into robust watermarking and fragile watermarking. Robust watermarking refers to the ability to withstand a large number of different physical and geometric distortions, including intentional and unintentional. If an attacker tries

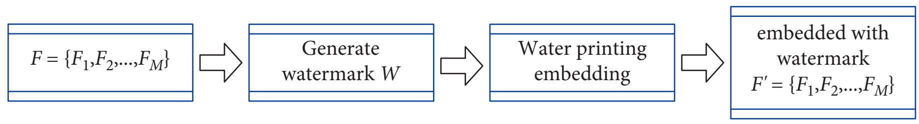


Figure 4: Flowchart of watermark embedding.

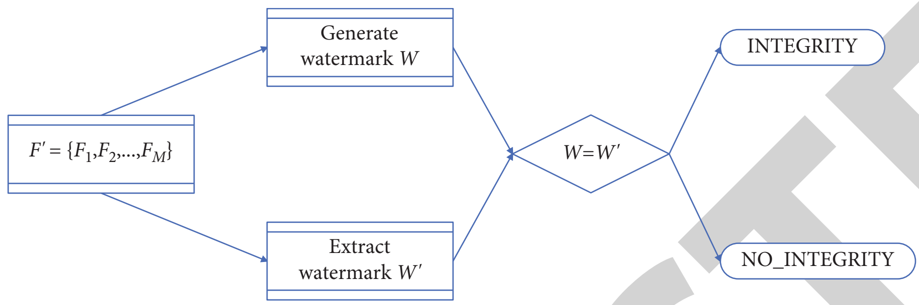


Figure 5: Flowchart of watermark detection.

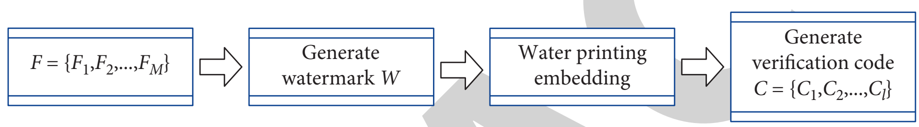


Figure 6: Flowchart of watermark embedding.

to delete the watermark, it will cause the complete destruction of digital products. The fragile watermark is to embed a digital watermark into the data under the premise of ensuring certain data quality. When the data content changes or is suspected, the watermark is extracted to identify the authenticity of the data content and point out the location of the tampering and changes, even tampering with the type, etc., so as to achieve accurate authentication of the integrity and authenticity of the data content. According to whether digital watermarks can be perceived, that is, whether they can be perceived by human vision, digital watermarks can be divided into perceptible watermarks and nonperceptible watermarks. Perceivable watermark means that the existence of a watermark can be observed, such as a logo in an image or video. The imperceptible watermark refers to the unwatermarked carrier data and the watermarked carrier data, which cannot be distinguished by human visual observation. According to whether there are private secret parameters when the watermark is embedded, digital watermarks can be divided into private watermarks and public watermarks. Private watermark means that the watermark is based on some private parameters (such as the key) during the embedding process, and only those with permission can verify the watermark information and prove the ownership of the data. However, in a private watermark, if the private parameters are leaked, the attacker can easily remove the watermark or add fake watermark information to the data through the private parameters. Public watermark means that any user can verify the watermark information multiple times to detect whether the data content has been tampered with, and this process does not require any information about private parameters.

Next, combined with the actual needs of digital protection of cultural heritage, this paper uses the vector data algorithm of GIS in the model construction.

Whether to generate watermark information, compare the generated watermark information with the extracted watermark information (or a public verification code). Then, to generate watermark information, compare the generated watermark information with the extracted watermark information (or a public verification code). If the two are the same, it is determined that the vector geographic data have not been tampered with. If the two are different, the tampered vertex position can be located according to the mismatched watermark. The specific process is shown in Figure 2.

In the vector data model of GIS, spatial entities are represented in the form of points, lines, and areas. The spatial relationship between these entities includes topological relationship, direction relationship, and metric relationship. Among them, the topological relationship is the key to spatial analysis. Topological relationship is used to describe the relationship between two objects. For example, there is a polygon on the rubber surface and a point inside the polygon. No matter if the rubber is compressed or stretched, the point still exists inside the polygon, and the spatial position relationship between the point and the polygon does not change, but the area of the polygon will change. The former is a topological relationship of space, but the latter is not a topological relationship. People generally locate a spatial target not by memorizing its spatial coordinates, but by determining the spatial position relationship between a target and other more familiar targets. This relationship is often a topological relationship, such as at which intersection or street a school is located.

The dimensionally extended nine-intersection model is a spatial model composed of the interior (A°), boundary (∂A), and exterior (A^{-}) of surface feature A and the interior (B°), boundary (∂B), and exterior (B^{-}) of B.

The dimensionally extended nine-intersection model formally describes the topological relationship of discrete spatial objects, and each element in the matrix has two values of "empty" and "nonempty." It can express a total of $2^9 = 512$ possible spatial relationships, but in reality, some relationships do not exist.

The topological relationship between any two surface features in the model is geometrically invariant, that is, translation, rotation, and zoom operations will not change the topological relationship between surface features.

The watermark information is generated according to the topological relationship between the features. The proposal of the watermarking algorithm requires the use of the concept of separation relationship measurement value. In order to calculate the separation degree between any two separated features, surface feature A is used as a reference object to measure the separation between A and B. The calculation method of the degree of separation is shown in the following formula:

$$EC(B) = \frac{\text{area}(B)}{\text{area}(B \oplus \text{dist}(\partial B, \partial A))}, \tag{1}$$

$\text{dist}(\partial B, \partial A)$ represents the shortest distance between A and B, $\oplus$ is the area expansion operator, and $\text{area}(B \oplus \text{dist}(\partial B, \partial A))$ represents the area after B is expanded by $\text{dist}(\partial B, \partial A)$. The schematic diagram of the calculation of the detachment relationship measure is shown in Figure 3.

The separation relationship measurement value is calculated from the shortest distance between two separated surface features and the area of the ground objects. The specific steps are as follows:

Step 1: the algorithm calculates the shortest distance between two surface features A and B. It is assumed that surface feature B needs to be embedded with a watermark, and surface feature A is the reference surface feature of surface feature B. If the shortest distance between surface feature B and surface feature A is 0, the algorithm replaces the reference surface feature.

Step 2: the algorithm expands the surface feature B to a position exactly tangent to the surface feature A.

Step 3: the algorithm compares the area before the expansion of the surface feature B to the area after the expansion, and the result of the ratio is the measurement value of the separation relationship between the separated surface features A and B.

A method for topological integrity verification of geographic data based on weak watermarking vectors is proposed. The watermark information is generated by the spatial separation distance between the features, and the ratio of the separation distance is modified, and the features are scaled according to the ratio to achieve the purpose of embedding the watermark. The watermark detection verifies

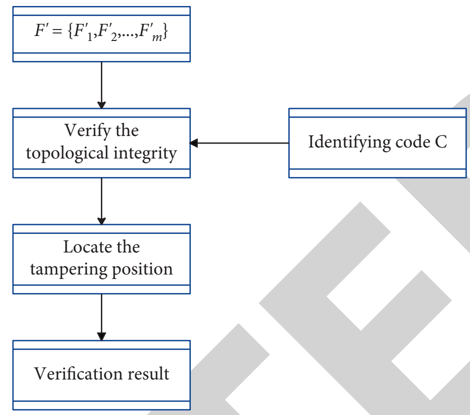


Figure 7: Flowchart of watermark detection.

the topological integrity of the vector geographic data set by judging whether the spatial topological relationship between the features has changed.

The topological integrity verification algorithm of vector geographic data is divided into two parts: watermark embedding algorithm and watermark detection algorithm. The watermark embedding algorithm first generates a watermark string according to the spatial topological relationship of the vector geographic data set, and then embeds the generated watermark string into the vector geographic data set. By slightly modifying the vector geographic data, it achieves the purpose of watermark embedding. The watermark embedding flowchart is shown in Figure 4.

The watermark detection algorithm is used to verify the topological integrity of the vector geographic data set. First, generate a watermark string from the spatial topological relationship of geographic data, and extract the original watermark string at the same time, and then compare the generated watermark string with the extracted original watermark string. If the two are the same, it indicates that the topological relationship of the vector geographic data set has not been destroyed. On the contrary, it indicates that the topological relationship of the vector geographic data set has been destroyed and its data have been tampered with. The specific watermark detection process is shown in Figure 5.

In a vector geographic data set, it usually contains many surface features, which we call features, and each feature is composed of many vertices. The set of surface features in the vector geographic data set is $F = \{F_1, F_2, \ldots, F_m\}$, and m represents the number of surface features. The set of vertices contained in each surface feature is denoted as $F_i = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, and n represents the number of vertices of the surface feature F_i.

Combined with the watermarking algorithm, the process of digital protection of cultural heritage is analyzed and identified.

Generate watermark:

(1) Calculate the shortest distance between surface features. The shortest distance between two features in the vector geographic data set is calculated. If the figure to be embedded with the watermark is B, and figure A is the reference figure of figure B, the

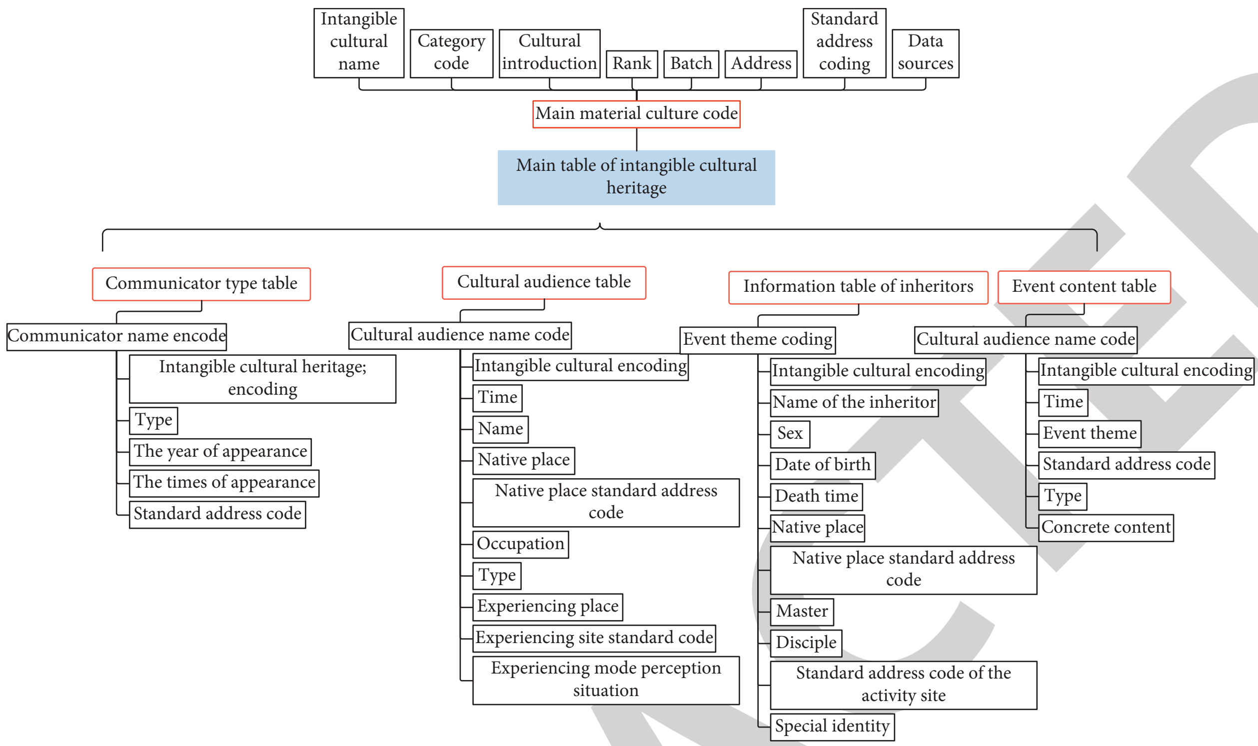


Figure 8: The structure of the attribute table centered on intangible culture.

distance between the two figures from the closest point is the shortest distance between figure A and figure B. The calculation method is as follows:

$$\operatorname{dist}(\partial B, \partial A) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}. \tag{2}$$

Among them, point (x_A, y_A) and point (x_B, y_B), respectively, represent the nearest vertex coordinates between surface feature A and surface feature B.

(2) Calculate the distance value of the spatial topological relationship. The algorithm calculates the spatial distance value between the surface feature B to be embedded in the watermark and the reference surface feature A, and the result is denoted by D. The calculation method is as follows:

$$D = \frac{\operatorname{area}(B)}{\operatorname{area}(B \oplus \operatorname{dist}(\partial B, \partial A))}. \tag{3}$$

Among them, $\operatorname{area}(B)$ is the area of surface feature B, $\operatorname{area}(B \oplus \operatorname{dist}(\partial B, \partial A))$ represents the area of surface feature B after expansion $\operatorname{dist}(\partial B, \partial A)$, and the set $D = \{D_1, D_2, \ldots, D_k\}$ is obtained, and k is the length of the watermark string.

(3) Generate watermark bits. The generation of each watermark bit is calculated by the distance D_i of the spatial topological relationship. According to the obtained result, the watermark bit w_i is generated, then the watermark set is $W = \{w_1, w_2, \ldots, w_k\}$.

$$w_i = \operatorname{Integer}(D_i) \bmod 2. \tag{4}$$

The algorithm modifies the least significant bit of the distance D_i according to the watermark bit w_i. When the watermark bit is 1, the algorithm modifies the value $\operatorname{LSB}(D_i)$ of the least significant bit of D_i to the range of 5–9. When the watermark bit is 0, the value f of the lowest bit of D_i is modified to the range of 0–4, as shown in formula (5). The surface feature B will be scaled according to the ratio s of the adjusted separation distance D_i' to D_i, so as to realize the embedding of the watermark.

$$\operatorname{LSB}(D_i) = \begin{cases} 5 \sim 9, & w_i = 1, \\ 0 \sim 4, & w_i = 0. \end{cases} \tag{5}$$

The fragile watermark method can simply and effectively detect the topological integrity of vector geographic data, but the only disadvantage is that the original data value needs to be modified during the watermark embedding process. When the data accuracy is very high, and the data user does not want to modify any original data value, the fragile watermark method cannot meet this demand. Therefore, in order to solve the above problems and to verify the integrity of the topological relationship of vector geographic data for many times, a vector geographic data topological integrity verification algorithm based on public watermark is proposed. The data user can verify the topological integrity of the vector geographic data through the obtained public verification code on the public platform.

The watermark embedding process is mainly divided into three steps: watermark generation, watermark embedding, and verification code generation. The watermark embedding process does not modify the original data value, and the watermark embedding flowchart is shown in Figure 6.

(1) Calculate the shortest distance between surface features. The algorithm assumes that the surface feature to be embedded in the watermark is B, and surface feature A is the reference surface feature of surface feature B. Then, the distance between the two surface features from the closest point is the shortest distance between surface feature A and surface feature B, and the calculation method is as follows:

$$\mathrm{dist}(\partial B\ ,\partial A) = \sqrt{\left(x_B - x_A\right)^2 + \left(y_B - y_A\right)^2}. \tag{6}$$

Among them, (x_B, y_B) and (x_A, y_A) represent the nearest vertex coordinates on surface feature B and surface feature A, respectively. If the value of $\mathrm{dist}(\partial B\ ,\partial A)$ is 0, then the surface feature B is matched with another surface feature. When the shortest distance between the two surface features is not 0, the matching is successful. The reference surface feature of surface feature B is no longer used as the reference surface feature of other surface features that need to be embedded with watermark, and will not be embedded with watermark information.

(2) Calculate the distance measurement value of spatial topological relationship. The algorithm calculates the spatial separation metric value between the surface feature B to be embedded in the watermark and the reference surface feature A, and it is denoted by D. The calculation method is as follows:

$$D = \frac{\mathrm{area}(B)}{\mathrm{area}(B \oplus \mathrm{dist}(\partial B, \partial A))}. \tag{7}$$

Among them, $\mathrm{area}(B)$ is the original area of surface feature B, and $\mathrm{area}(B \oplus \mathrm{dist}(\partial B, \partial A))$ is the area after B is expanded by $\mathrm{dist}(\partial B, \partial A)$. After calculating the distance measurement values of all surface features in the geographic data set, a set of distance measurement values $D = \{D_1, D_2, \ldots, D_l\}$ is obtained, where l is 1/2 of the total number of surface features in the geographic data set.

(3) Generate watermark. The generation of the watermark is calculated by the distance D_i of the spatial topological relationship. The algorithm hashes D_i into a value W_i of length n, $W_i = \{w_1, w_2, \ldots, w_n\}$ is the watermark information of the current surface feature, and n is the total number of vertices contained in the embedded watermark surface feature.

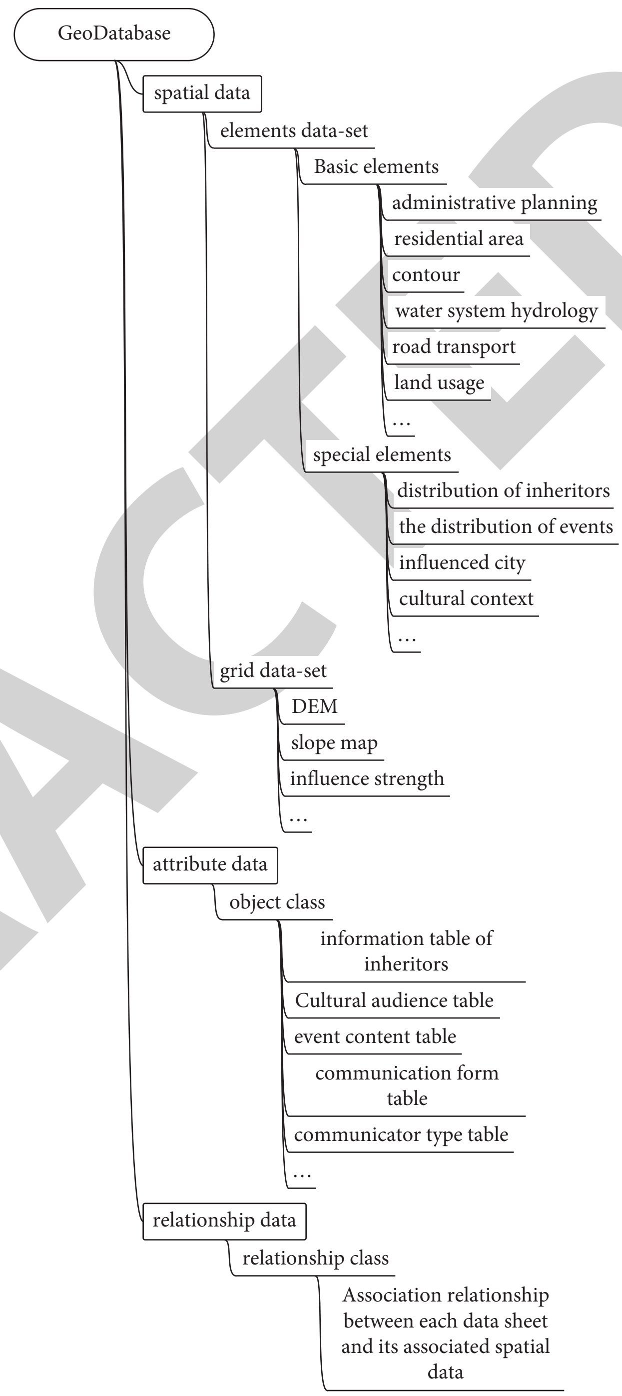


Figure 9: Database model.

$$W_i = \mathrm{Hash}(D_i, n). \tag{8}$$

First, the algorithm modulates the coordinate value of each vertex of the embedded watermark surface feature to 2 to obtain the bit position $V_i = \{v_1, v_2, \ldots, v_n\}$ of each vertex of the surface feature. Then, the algorithm performs an XOR operation on each bit of the watermark bit W_i and each bit of V_i to obtain the surface feature verification code c_i. The calculation method is as follows:

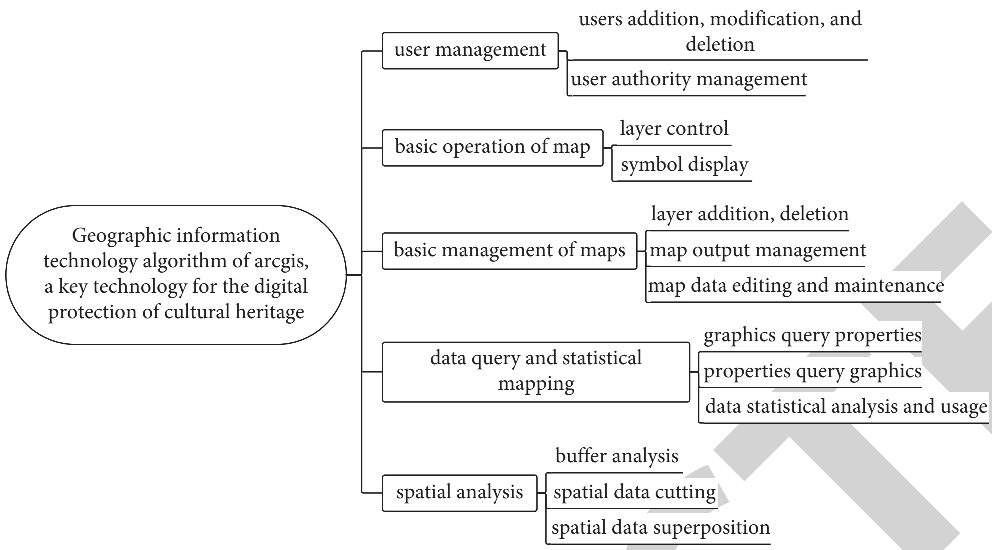


Figure 10: The overall function module of the system.

$$W_i \otimes V_i = c_i. \tag{9}$$

In the watermark detection, the calculated spatial separation relationship distance value is first compared with the public verification code, and the topological integrity of the vector geographic data set F'' is determined according to the comparison result. If the spatial topological relationship of the features is destroyed, the unmatched verification code can be used to locate the tampered location. The watermark detection flowchart is shown in Figure 7.

The algorithm verifies the topological integrity.

(1) Calculate the shortest distance between surface features. The shortest distance between the two surface features in the vector geographic data set is calculated. The calculation method is the same as formula (6) in the watermark embedding.

(2) Calculate the distance measurement value of spatial topological relationship. The algorithm calculates the spatial distance between the surface feature B and the reference surface feature A to be watermarked, denoted by D', and the calculation method is as follows:

$$D' = \frac{\text{area}\,(B)}{\text{area}\,(B \oplus \text{dist}\,(\partial B, \partial A))}. \tag{10}$$

The algorithm obtains the set $D' = \{D'_1, D'_2, \ldots, D'_l\}$ of distance values of the vector geographic data set.

(3) Verify topology integrity. The algorithm extracts the D_i in the verification code C_i, and compares the extracted D_i with the generated D_i. If D_i and C_i are the same, it means that the surface feature has not been tampered with. If they are different, it means that the surface feature has been tampered with, and then the position of the tampered vertex needs to be further found.

Determine the location of tampering:

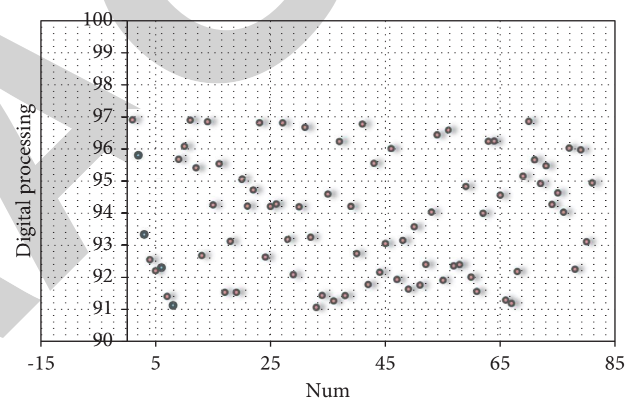


Figure 11: Statistical diagram of the effects of key technologies in the digital protection of cultural heritage.

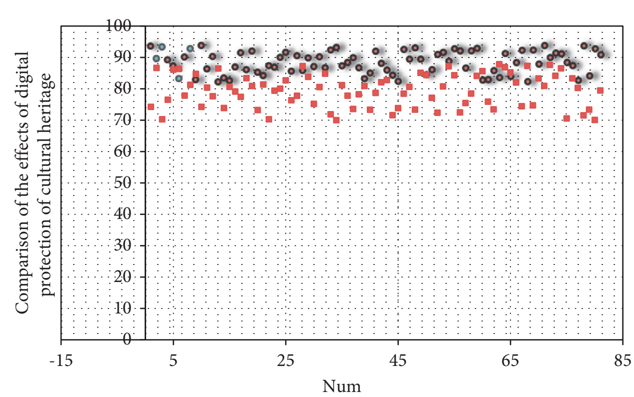


Figure 12: Comparison of algorithm effects.

(1) Restore watermark information. The algorithm extracts c_i in the verification code C_i, and hashes D_i according to the length n of c_i to obtain the original watermark information $W_i = \{w_1, w_2, \ldots, w_n\}$.

$$W_i = \text{Hash}\,(D_i, n). \tag{11}$$

Table 1: Method comparison results.

Number	The method of this paper	The method of Yakar and Doğan [6]	Number	The method of this paper	The method of Yakar and Doğan [6]
1	86.41	64.19	25	76.98	69.45
2	70.72	68.33	26	75.01	58.38
3	78.55	59.61	27	79.64	70.18
4	82.44	58.21	28	86.96	70.79
5	76.81	62.18	29	85.76	69.49
6	76.32	58.85	30	78.93	68.65
7	85.81	69.16	31	84.14	63.17
8	79.78	60.04	32	73.37	64.17
9	79.93	70.82	33	74.93	59.65
10	86.01	63.41	34	83.44	69.73
11	72.46	58.53	35	72.52	63.27
12	77.06	59.44	36	77.51	64.60
13	79.97	67.34	37	86.80	68.13
14	70.20	63.01	38	83.73	64.46
15	80.11	69.32	39	72.03	63.89
16	84.13	58.43	40	73.35	65.61
17	77.99	59.87	41	79.97	60.16
18	83.88	61.04	42	75.82	68.20
19	79.19	66.68	43	81.47	64.96
20	78.68	63.98	44	76.39	64.98
21	80.06	61.24	45	85.25	63.83
22	79.66	63.49	46	79.81	70.32
23	86.74	58.38	47	86.15	59.76
24	72.38	59.88	48	84.09	62.23

(2) Determine the location of tampering. The algorithm performs XOR on the extracted original watermark information W_i and c_i to obtain all the original vertex values V_i of the surface feature. At the same time, the algorithm modulates all vertex coordinates of the current surface feature with 2 to generate $V_i' = \{v_i', c_2', \ldots, v_n'\}$. Then, the algorithm compares V_i and V_i'. If V_i and V_i' are completely equal, the current surface feature has not been tampered with; otherwise, the tampered vertices can be found by filtering the vectors where V_i and V_i' are not equal.

4. Digital Protection Technology of Cultural Heritage Based on ArcGIS Geographic Information Technology Algorithm Model

When conducting a special research on a certain intangible culture, we first search for a certain cultural name according to the intangible cultural code, and match the corresponding spatial location data and event content according to the standard address code and cultural introduction in the search results. Moreover, we query the detailed content of the event according to the event subject code in the event content table, and learn the basic information of the inheritor and the scope of space activities by querying the inheritor code. In addition, we use the communicator type code to connect the inheritor information table and the event content table, etc., as shown in Figure 8.

This article uses ArcGIS' geographic information technology algorithm and Geo Database to construct a spatiotemporal database. The database includes three parts, namely, spatial data, attribute data, and relational data. The basic structure of the spatiotemporal database is shown in Figure 9. Among them, the attribute data include the inheritor information table, cultural audience table, event content table, dissemination form table, and communicator-type table integrated above. The relationship data defines the connection fields and association relationships between the tables, and is associated with a unique identification code.

The overall function module diagram of the system is shown in Figure 10.

After constructing the above model, the effect of the system model constructed in this paper is verified, the digital processing of cultural heritage is verified, and the digital verification results are counted, and the results shown in Figure 11 are obtained.

As can be seen from the above figure, the key technologies for the digital protection of cultural heritage proposed in this paper have good effects in the digital processing of cultural heritage. After that, the performance of the algorithm proposed in this paper is verified, and the method proposed in this paper is compared with the traditional method, and the result shown in Figure 12 is obtained.

The research content of this paper is compared with the method proposed in the literature [6], and the results are shown in Table 1.

From the above research, we can see that the digital protection technology of cultural heritage based on ArcGIS geographic information technology algorithm in this paper can play an important role in the digital protection of cultural heritage.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 9761517, 1 page
https://doi.org/10.1155/2022/9761517

Retraction

Retracted: A Graph Neural Network (GNN) Algorithm for Constructing the Evolution Process of Rural Settlement Morphology

Security and Communication Networks

Received 9 November 2022; Accepted 9 November 2022; Published 22 November 2022

Security and Communication Networks has retracted the article titled "A Graph Neural Network (GNN) Algorithm for Constructing the Evolution Process of Rural Settlement Morphology" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the Chief Editor.

References

[1] Z. Hu, K. Chen, and X. Xie, "A Graph Neural Network (GNN) Algorithm for Constructing the Evolution Process of Rural Settlement Morphology," *Security and Communication Networks*, vol. 2022, Article ID 7517347, 10 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2020, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7517347, 10 pages
https://doi.org/10.1155/2022/7517347

Research Article

A Graph Neural Network (GNN) Algorithm for Constructing the Evolution Process of Rural Settlement Morphology

Zhe Hu[1], **Kexin Chen**[1,2] **and Xiaofei Xie**[3]

[1]*School of Architecture and Urban Planning, Huazhong University of Science and Technology, Wuhan City 430074, China*
[2]*School of Civil Engineering and Architecture, Wuhan Institute of Technology, Wuhan City 430205, China*
[3]*School of Landscape Architecture, Huaihua University, Huaihua, Hunan 418000, China*

Correspondence should be addressed to Zhe Hu; 2014210046@hust.edu.cn and Xiaofei Xie; xxf@hhtc.edu.cn

Received 31 December 2021; Revised 19 January 2022; Accepted 26 January 2022; Published 4 March 2022

Academic Editor: Chin-Ling Chen

Traditional statistical methods were mainly used to study the evolution process of rural settlement form and scale from a qualitative perspective, but it was difficult to quantitatively analyze the evolution process of the rural settlement form. Therefore, this paper proposed an intelligent monitoring method of rural settlement morphology evolution process based on the graph neural network (GNN) algorithm. Firstly, the specific working process of image feature extraction, analysis, and processing based on the graph neural network (GNN) algorithm was described. Secondly, combined with the change characteristics of rural settlement morphology evolution and scale development, the graphical neural network algorithm was used to effectively extract the morphological characteristics of rural settlements, and the monitoring information characterizing the dynamic changes of rural settlement morphology and scale was obtained through feature clustering. Finally, through experiments and using the graph neural network algorithm, the evolution process of rural settlement morphology was monitored in real time. The experimental results showed that the monitoring data obtained by this method were basically consistent with the actual statistical results, which showed that the intelligent monitoring method of the rural settlement form evolution process based on graph neural network algorithm can better reflect the dynamic change process of the rural settlement form and scale development. This study will provide some theoretical reference and guiding significance for the quantitative analysis of the evolution process of the rural settlement morphology and its influencing factors.

1. Introduction

With the acceleration of urbanization, some rural settlements have developed rapidly; especially, the form, function, and scale of settlements have been greatly developed. Compared with plain areas and traditional agricultural areas, the shape and distribution of rural settlements in mountainous areas are complex and changeable [1]. Therefore, it is of certain guiding significance for the rational guidance, regulation, and optimization of the scale of rural settlements to study the change process of temporal and spatial characteristics of rural settlements in mountainous areas and deeply explore the evolution law and influencing factors of rural settlements.

According to the theory of geography, rural settlements are important places for farmers' production, life, and socialization. From the perspective of geographical composition and morphological changes, rural settlements are patches of interaction between farmers and land. Therefore, at rural settlements, people are interdependent and have certain regional structural characteristics and functions. The morphological evolution of rural settlements is affected by the conditions of natural resources and the level of economic and social development [2]. There are some differences in the morphological evolution characteristics, speed, and process of different rural settlements. Therefore, through the exploration of the evolution process of the rural settlement form, scholars at home and abroad have revealed the relationship between the evolution of rural settlement spatial pattern and different regions and development backgrounds, which is also one of the research hotspots of rural settlement.

The form and functional structure of rural settlements are dynamic, sustainable, and long-term, and their evolution process is a complex and changeable process. Many years ago, some scholars conducted an in-depth research on the relationship between the temporal and spatial structure of rural settlements and the external environment [3, 4]. With the deepening of rural settlement research, the research on the evolution of rural settlement morphology and structure has been extended to population migration, urban-rural integration, land development, and cultural life. In the 21st century, with the improvement of rural settlement research methods, many scholars turn their research focus to the evolution mechanism and prediction of rural settlement morphology. The research shows that the continuous development of the rural settlement form is closely related to the external natural environment and social stage. The development scale, form, and structure of the rural settlement can fully reflect the internal relationship between residents' life and nature. In addition, from the perspective of landscape ecology of rural settlements, some scholars use the change of landscape index to analyze the landscape pattern and dynamic change of rural settlements, so as to reveal the development process of the evolution of rural settlements. Traditional methods are usually used to explore the development process of rural settlement evolution from a qualitative perspective, but cannot quantitatively reflect the rural settlement form and its dynamic changes.

2. Related Works

As early as the 19th century, people began to study rural settlements and their functional structure. In 1841, German scholars analyzed the reasons for the formation and development of rural settlements, which provided a theoretical basis for the later study on the evolution of settlement form. Forman put forward the central geography theory for the change of rural settlements, and expounded that the scale and distribution of rural settlements are related to economic, transportation, and administrative factors [5]. French scholar Blanches and Wesolowska expounded the correlation between the formation of rural settlements and natural environmental factors and historical and cultural factors, and found that the settlement form and its development change with the change of geographical location [6]. According to the types of rural settlements in different regions of Germany, German scholars put forward the basic theories and methods of settlement geography. With the extensive study of settlement area and morphology, many scholars have carried out an in-depth research on rural settlements from different aspects and made some progress.

From the research status and future development trend of rural geography, some scholars in the United States and Britain analyzed the geography, history, and their relationship with rural settlement areas, and explored the relationship between rural settlement morphological structure adjustment, economic and social development, and landscape change in the industrial age [7, 8]. In addition, some scholars have studied the relationship between the morphological structure of rural settlements and other adjacent disciplines. The scale and evolution of rural settlements can usually be analyzed and extracted by means of system analysis, quantitative and deductive analysis, combined with mathematical model and geographic information system, to reproduce the evolution process of rural settlements in an intuitive way.

In recent years, 3S technology has been gradually applied to the study of rural settlements. For example, GIS technology can effectively obtain various remote sensing image information such as geographical and spatial location. In addition, the quantitative analysis method based on GIS technology provides an effective means for the study of rural settlements. For the research on the evolution of rural settlement spatial form and its influencing factors, the early stage mainly explored the influence law of natural geographical factors on the distribution of rural settlements based on the results of field investigation and analysis and from a qualitative point of view [9, 10]. With the increasing development of geographic information system and 3S technology, the way of data acquisition has been improved, and the accuracy of data obtained has been improved. People began to use the method of landscape ecology to study the pattern characteristics of rural settlements and the evolution characteristics of settlement patches. At the same time, mathematical statistics was used to study the influence of natural and human factors on settlement evolution. Since then, people's research on rural settlements has changed from qualitative description to quantitative analysis.

In exploring the evolution process of rural settlement morphology, most of the existing methods only process the image appearance features obtained from remote sensing images, and these features usually lack in-depth information, which makes it difficult to obtain ideal prediction results when processing complex image information [11, 12]. In recent years, depth neural network model has been greatly popularized in the field of vision. Among them, some scholars have studied the image segmentation method based on depth learning model, which can be effectively applied to the processing and analysis of remote sensing images. Inspired by the research results in other related fields, some scholars apply graph neural network model to image feature extraction and processing, and combine semantic enhancement and network segmentation methods to express different features. Using the attention mapping mechanism, the image features at different positions can be weighted to obtain the feature map, but this method requires more computing resources and storage space [13]. The method of graph neural network is usually to segment the collected image to form multiple small regions, then use the traversal method to reorganize the two regions with the largest correlation, and re aggregate the image into one region until all relevant pixels are concentrated in one candidate region. Then, the image feature points are selected from the candidate regions, stretched, and enlarged into pictures of the same size, and the pictures are sent to CNN model for processing to obtain the required image feature points.

Compared with other image processing methods, the method based on graph neural network can subdivide and fuse the image. Therefore, it can be better applied to the real-time dynamic analysis of the image.

3. Working Principle of Graphical Neural Network

3.1. Structure and Function of Graph Neural Network. According to the application requirements of structured and unstructured scenes, some researchers proposed using graph neural network (GNN) to process the structure information of graphs and achieved good results. GNN network can not only realize the problems that are difficult to be solved by other neural network methods but also make the deep learning theory widely used in the fields of recommendation system, graph clustering, and so on. Compared with other network structures, GNN network structure is simpler [14]. It mainly adopts graph convolution and graph pooling to complete relevant operations, and uses the full connection layer and output layer to form the whole network structure, as shown in Figure 1.

When the GNN network structure is applied to image processing, the initial image matrix is generally transmitted to the input layer as the input object of the GNN network structure, and then the graph convolution, graph pooling, and other operation modules are used for correlation processing. Finally, the processing results are output through the output layer. The main task of graph convolution is to linearly combine each feature point in the graph with its adjacent feature points, use the adjacency matrix to process the relevant feature information and propagate the results to different network layers. Using nonlinear layer to transform different morphological feature points, similar morphological feature points can be associated.

In order to realize the mapping of different images, the feature extraction can be carried out through the graph volume kernel, and the pooled method can be used to reduce the dimension of the extracted features. Therefore, multi-layer clustering algorithm can be used to reduce the dimension, which is also called graph pooling. The data types processed by the pooling layer are generally regular morphological features. In order to make the features extracted from any region applicable to other different regions, the features of different local regions can be aggregated and represented by larger dimensional morphological features. For example, in convolutional neural networks, maximum pooling algorithm and average pooling algorithm are often used to obtain the feature sizes of different regions, respectively, and expressed by the maximum or average value of the image features of the region.

The graph convolution operation of graph neural network is used to process the image features and output the results. The image features processed by the graph pooling layer are in the same neighborhood, and the feature points in this field form a cluster. Because the output results generated by graph convolution operation will cluster the feature points in the same field, graph neural network uses graph convolution operation to generate many graph features, and outputs graph feature information of different dimensions through feature clustering operation. Therefore, clustering algorithm can be used to process the features, so that the output feature dimensions are different.

In order to speed up the processing speed of the graph pooling layer, when the graph pooling operation is carried out on the graph features, it is necessary to use the adjacency matrix to calculate the graph feature pixels, and make the graph pooling layer and the upper layer connect orderly in the graph feature processing [15]. Among them, the discontinuous feature points formed during the graph pooling operation on the graph features do not affect the pooling output results. The pool operation process is shown in Figure 2.

3.2. Graph Neural Network Method. From the above structure and function of graph neural network, it is known that when using the graph neural network model to process image features, it is mainly to optimize the convolution kernel parameters of the network model, and make it reach the best value through repeated intensive training. Therefore, in this paper, polynomial expansion is applied to convolution kernel operation. When strengthening network model training, the optimization of convolution kernel parameters is transformed into continuous optimization of polynomial coefficients. Therefore, each optimization of the polynomial coefficients is based on the output of the previous model training [16]. The loss function used in this paper as the benchmark for training and evaluation of network model parameters is expressed as follows:

$$l = -\frac{1}{m}\sum_{x} y_x \ln b_x. \tag{1}$$

The above function can represent the relationship between the output result b of the input image sample x processed by the graph neural network model, the predicted image y, and the number of samples. When the error between the output result and the actual value is large, the loss function can be used and the parameters of graph neural network can be optimized quickly through the intensive training of the model. The GNN network model algorithm used in this paper mainly includes two parts: forward processing and backward processing. The image features are processed forward based on the GNN network model. The purpose is to process the image structure features by Fourier transform, and then take them as the input object of image convolution for subsequent processing [14]. The processing function of image features in the image convolution layer can be expressed as follows:

$$z^2 = \lambda\left(y_{x,i} + a\right) = \lambda\left(\sum_{k}^{j} g\omega_{k,i}(l) z_{x,i} + a\right), \tag{2}$$

where z^2 represents the output result of the image feature after the image convolution operation, λ is the activation function used by the image convolution, $y_{x,i}$ represents the output result of the image feature after the image convolution operation, and a is the constant parameter of the

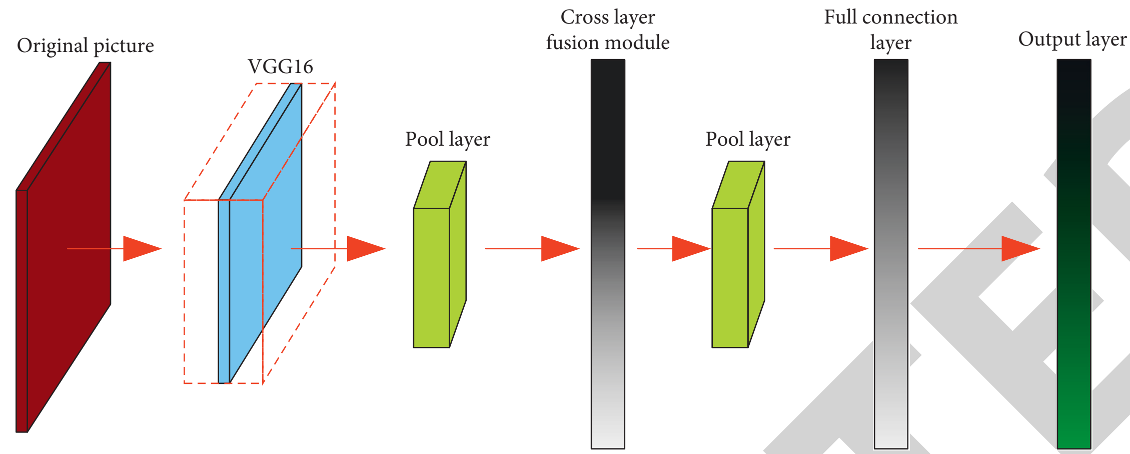


FIGURE 1: Structure diagram of graph neural network.

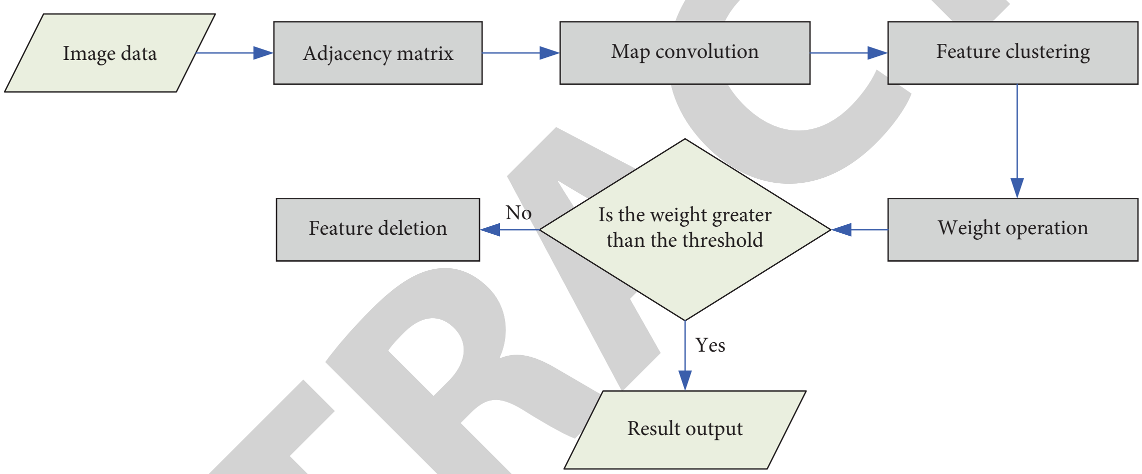


FIGURE 2: Diagram of pool layer operation process.

image convolution. Using the graph pooling layer of graph neural network model for graph pooling operation is mainly to cluster the feature points of the input image and transform them into one-dimensional feature images. The graph convolution processing in the network model is to input the one-dimensional graph features after the graph pooling operation to other layers, to further complete the graph feature extraction operation. By reducing all graphic features and transforming them into one-dimensional graphic feature vectors, one-dimensional graphic feature vectors are spliced and integrated as the input object of the whole connection layer of the graphic neural network model [16, 17]. The expression of splicing and integration of graph feature vectors in the full connection layer of graph neural network model is as follows:

$$z^{ih} = \lambda\left(q^{ih} z^{2c} + a^{ih}\right), \tag{3}$$

where λ is the activation function used by the full connection layer, z^{2c} represents the image feature processing result of the image pooling layer and takes it as the input processing object of the full connection layer, z^{c2} represents the weight coefficient of the image pooling layer, z^{c2} is the constant parameter of the layer, and z^{ih} represents the image feature output result after splicing and integration. The relevant model parameters are modified by activation function, and the output results are obtained by graph neural network model processing [18]. The forward processing algorithm flow is shown in Figure 3.

For the reverse processing algorithm in graph neural network, the best model parameters and weights are obtained mainly through the learning, training, and optimization of network model. In the graph neural network model, the forward processing algorithm is used to obtain the loss function, and then the loss function is used to calculate the error value, which is fed back from the lower layer of the network model to the upper layer in turn. Finally, the gradient method is used to optimize the parameters of the graph convolution.

Taking the collected image samples as the input object of the model, the network model is used to preprocess the input image. According to each pixel of the image, a graphic feature adjacency matrix is established to make the output of the model closer to the prediction target. Then, the graph features of graph convolution operation are checked by graph convolution. After the response processing of the

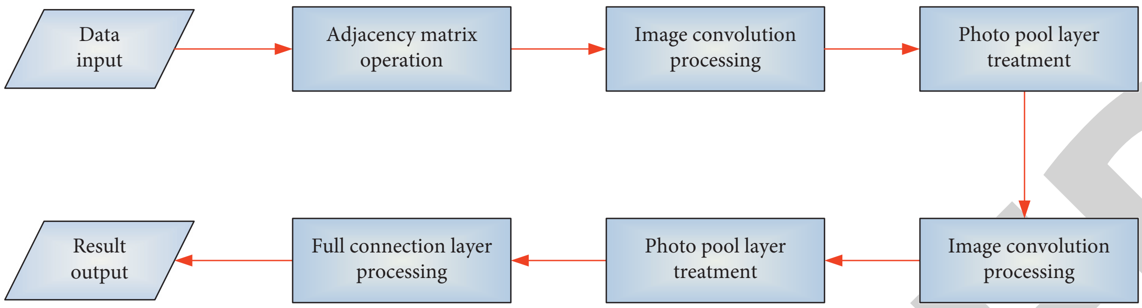


Figure 3: Flow diagram of the forward processing algorithm.

activation function, the graph feature adjacency matrix is used to aggregate different graph feature pixels. By aggregating the adjacent features of the current feature points, all feature pixels and their adjacency matrixes are continuously updated, to reflect the relationship between all graph features and their pixels. Finally, the clustering algorithm is used to classify the relevant features in the graph pool layer.

The model training mainly uses the experimental test results to verify the GNN network model, and the forward processing algorithm is used to obtain the output results. Because the algorithm trains and learns many sample images through GNN network model, and repeatedly uses the new model to forward process the image samples of the test set until the ideal prediction image is output, the continuous optimization of graph neural network model is the guarantee of outputting the ideal result [19]. In order to make the output of the model close to the real image samples, GNN reverse processing algorithm can be used and network training standards can be set. If the error obtained by the loss function cannot reach the preset standard, the error value needs to be continuously fed back to the network model, and then the network model parameters are adjusted through repeated training until the error obtained by the loss function meets the requirements, the network will not be trained and learned, and the finally trained network model will be used for image processing. The training process of graph neural network model is shown in Figure 4.

4. Analysis of Rural Settlement Form and its Scale Change

4.1. Analysis on the Change of Rural Settlement Form. Taking the typical rural mountainous areas with diverse topographic characteristics as the research object, this paper analyzes the morphological characteristics and evolution law of rural settlements in mountainous and hilly areas by using the methods of spatial analysis and econometrics. According to the spatiotemporal evolution characteristics of rural settlements, this paper mainly uses models such as land change index, nearest neighbor index, and nuclear density index to analyze the spatiotemporal change characteristics such as land scale and morphological distribution in the evolution process of rural settlements in this mountainous area, and combines the relevant characteristics from multiple angles in order to construct the evolution process of rural settlement morphology [8, 9]. Land change index can better describe the change speed of land types and the evolution law of regional characteristics in rural settlements, and can be used to reflect the law of land use and change in this area. The land change index of rural settlement area can be expressed by the following formula:

$$M = \frac{R_t - R_e}{R_e} \times \frac{1}{u}, \tag{4}$$

where R_e and R_t represent the total scale of rural settlements in the initial stage and later stage in turn, u is the time length of rural settlements, and M is the scale of rural settlements. $M < 0$ indicates that the residential land in the rural settlement area has been transformed into other land, reflecting that the scale of rural settlement has decreased, while $M > 0$ indicates that other land has been transformed into rural settlement land, reflecting that the scale of rural settlement has increased.

In order to analyze the gap of land change in rural settlements between different regions, it can be expressed by the relative land change rate, and its calculation formula is as follows:

$$r = \frac{\left|R_t - R_e\right|}{R_t} \times \frac{S_d}{\left|S_p - S_d\right|}, \tag{5}$$

where R_e and R_t represent the scale of rural settlements in the initial and final stages in turn, r and S_p represent the total scale of rural settlements in the initial and final stages, r represents the relative change rate of rural settlement land, $r > 1$ represents that the change rate of local area is greater than that of the whole; otherwise, it represents that the change rate of local area is smaller than that of the whole.

The spatial distribution of the morphological characteristics of rural settlements can be described by the nearest neighbor index, which is expressed by the ratio of the observed value to the expected value of the distance between adjacent points [20]. It can reflect the mutual proximity between different patches of rural settlements, which is expressed as follows:

$$K = \frac{H_v}{H_u} = \frac{\sum_{i=1}^{n} g_i/n}{\sqrt{n/S}} = \frac{\sqrt{\omega}}{n} \sum_{i}^{n} g_i, \tag{6}$$

where S represents the distance between point i and its nearest neighbor in the rural settlement area, H_v represents the average nearest neighbor distance between different

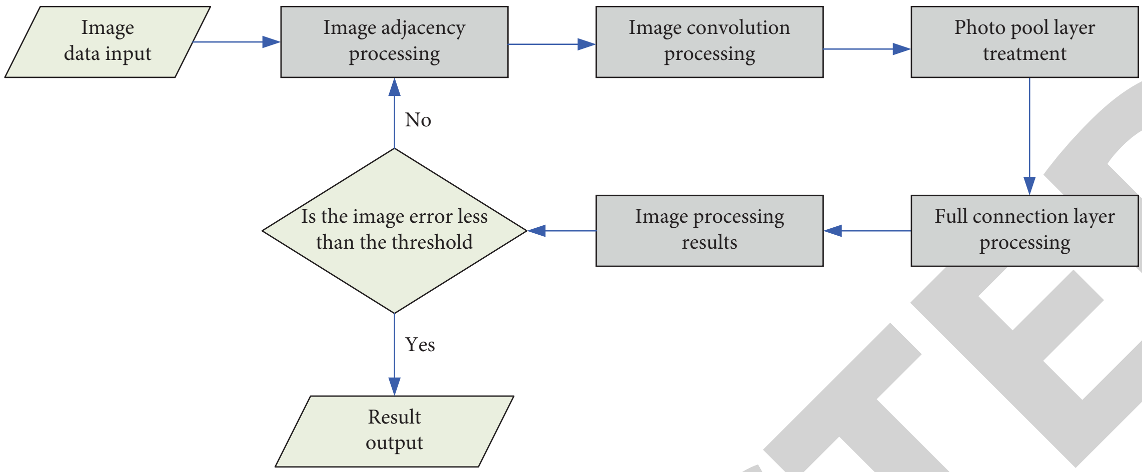


Figure 4: GNN network training update flow chart.

points, S represents the expected distance, and n represents the number of patches in the rural settlement, S represents the distribution density of each point in the rural settlement area, and S represents the minimum external rectangular area of different rural settlements. K is the nearest neighbor index. $K < 1$ indicates that the settlement patches are clustered, and $K > 1$ indicates that the settlement patches are randomly distributed.

Nuclear density index can reflect the morphological distribution characteristics of rural settlements and describe the spatial distribution of rural settlements [3]. Its calculation formula is as follows:

$$F(x_i, y_i) = \frac{1}{nd^2} \sum_{i=1}^{n} E\left(\frac{g_i}{n}\right), \tag{7}$$

where $F(x_i, y_i)$ is the predicted value of kernel density at a certain point of rural settlement at (x_i, y_i), g_i is the number of samples at different points of rural settlement, d is a constant, E is the kernel function, and $F(x_i, y_i)$ is the distance from point (x_i, y_i) to point i.

4.2. Analysis on the Change of Rural Settlement Scale. According to the existing research, the scale of rural settlements is a complex dynamic change process, mainly including the changes of the scale and morphological characteristics of rural settlements, as well as the changes of the regional structure and function of rural settlements. The evolution of rural settlement scale is not only related to the local economic and social development level but also changes with the changes of rural settlement structure and morphological characteristics, and leads to corresponding changes in rural settlement functions and services.

In the evolution process of rural settlement scale, the settlement function shows a certain law with the development of rural economy and society. It usually changes from homogeneous isomorphism to heterogeneous diversity, that is, from the original living farming state to the type of composite function. Because there are many types of rural settlements, the scale and evolution speed of rural settlements are related to the types of rural settlements. The traditional rural settlement land is mainly homestead, and the settlement function is usually small-scale cultivated land, enclosure, or breeding land around the homestead [5, 6]. With the rapid development of the rural economy and society, the scale of rural settlements is expanding, and the functions of settlements are also diversified. For example, the land for handicraft, commerce, and storage in the settlement area is gradually increasing. Due to the influence of various natural resources, and economic, social, and cultural factors, the scale, morphological structure, and function of rural settlements are constantly changing. The complexity, diversity, and dynamics of rural settlement functions are the main characteristics of rural settlement evolution in recent years.

4.3. Morphological Structure and Distribution Characteristics of Rural Settlements. The distribution of rural settlements is not only an important part of the evolution of rural settlements but also the main object of studying the changes of regional morphological characteristics of rural settlements [10]. The evolution characteristics of rural settlements mainly reflect the relationship between the morphological structure and function of rural settlements and time from the aspects of regional land type, form, scale, and agglomeration degree. The type and distribution of rural settlement land can reflect the change of regional land use. The change of rural settlement land use type and its area can reflect the change law of rural settlement form in land use.

Based on the statistical analysis of the land use data of a rural settlement area in 2000, 2010, and 2020, through the data and research, it is known that the land-use types of the rural settlement include forest land, cultivated land, water area, grassland, urban land, and rural settlement land. Figure 5 and Table 1 show the land-use types and land-use change results of the rural settlement area in 2000, 2010, and 2020.

According to the statistical results in Table 1, the land type of the rural settlement area is mainly forest land, followed by grassland, and cultivated land, indicating that the

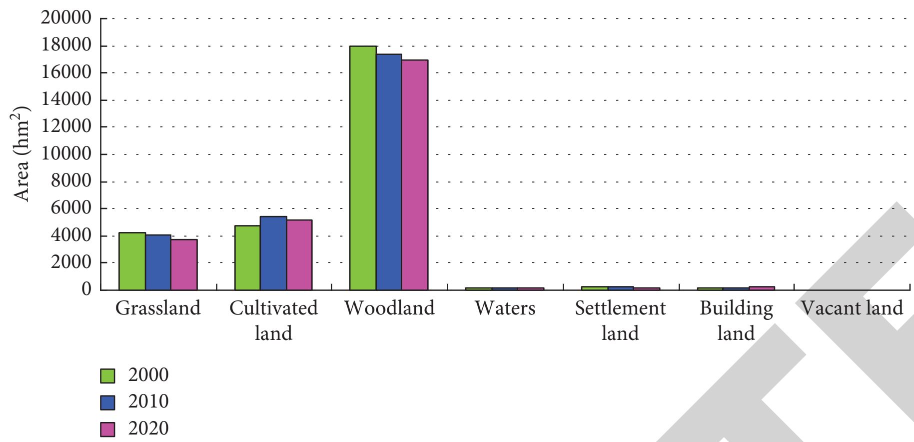


Figure 5: Statistics of land types in settlement areas from 2000 to 2020.

Table 1: Dynamic changes of various types of land use in rural settlements from 2000 to 2020.

Item	Rate of change (%)	Transfer out rate (%)	Transfer in rate (%)
Cultivated land	10.13	0.35	10.48
Woodland	−0.14	1.79	1.65
Grassland	−12.47	13.28	0.81
Waters	4.25	0.83	5.08
Settlement land	−1.29	1.64	0.35
Building land	8.36	8.29	16.65
Vacant land	−37.61	38.27	0.66

natural ecological environment of the settlement is good, the vegetation density is large, and the degree of land development is not high. Over time, from 2000 to 2020, the occupied area of grassland and forest land decreased, the water area was relatively stable, and the cultivated land area still showed a slight growth trend, but the construction land has been increasing significantly.

4.4. Analysis on the Morphological Characteristics of Rural Settlements. Studies have shown that rural settlements can be divided into four categories: living alone, small, medium, and large according to different patch forms. In order to describe the changes of rural settlement scale and its morphological characteristics, the patch number, patch proportion, patch area size, total patch proportion, average patch area, and other parameters in rural settlement area are selected for statistics, as shown in Tables 2 and 3.

The statistical results in Table 2 reflected that the morphological characteristics of rural settlements in 2000 were mainly medium-sized and large patches, which were 52.29% and 45.96%, respectively. Among them, the total area and proportion of medium patches were lower than large patches, the proportion of large patches was 75.55%, and the average area of patches reached 10.47 hm^2, which was higher than other settlement types. The number and proportion of patches in solitary and small settlements were 0.45% and 1.31%, respectively, and the total area and proportion of patches in solitary and small settlements were only 0.01% and 0.19%, respectively. The above analysis shows that in 2000, the distribution types of large settlements accounted for a large proportion, and the number of small settlements was small, which shows that the scale of rural settlements is significantly different. On the other hand, the statistical results in Table 3 reflected that the change of the scale of the rural settlement was not obvious from 2000 to 2010. The patch index of large-scale settlements decreased compared with 2000, indicating that the patch area of large-scale settlements had little impact on the change of the rural settlement form. The number of living alone patches increased to a certain extent, the number of small patches remained basically unchanged, and the number and total area of medium-sized and large patches were lower than those in 2000. This shows that the change of rural settlement form in 2010 is not significant compared with 2000.

5. Dynamic Monitoring of Morphological Changes of Rural Settlements based on Graph Neural Network

5.1. Monitoring Method of Rural Settlement Morphology based on Graph Neural Network. From the abovementioned changes in the morphological structure and function of rural settlements, in order to explore the evolution process of rural settlements, the graph neural network (GNN) algorithm can be used to dynamically monitor the change process of rural settlements. From the above analysis of the morphological characteristics of rural settlements, it is known that the morphological image of rural settlements has local characteristics, so it belongs to the local map structure. Because

Table 2: Statistical table of rural settlement scale in 2000.

Type	Area scope (hm^2)	Patch number	Patch proportion (%)	Patch area size (hm^2)	Total patch proportion (%)	Average patch area size (hm^2)
Alone	<0.2	23	0.45	2.07	0.01	0.09
Small	0.2–2	67	1.31	61.24	0.19	0.91
Medium	2–5	2684	52.29	7926.53	24.25	2.95
Large	>5	2359	45.96	24692.51	75.55	10.47

Table 3: Statistical table of rural settlement scale in 2010.

Type	Area scope (hm^2)	Patch number	Patch proportion (%)	Patch area size (hm^2)	Total patch proportion (%)	Average patch area size (hm^2)
Alone	<0.2	28	0.55	2.52	0.01	0.09
Small	0.2–2	68	1.34	62.35	0.19	0.92
Medium	2–5	2627	51.87	7479.24	23.36	2.85
Large	>5	2342	46.24	24468.65	76.43	10.45

the morphological structure of the local graph can obtain the relationship between the layers of the neural network by using the local mean operation method, the morphological characteristics of rural settlements can be used as the input characteristic graph and processed by the local mean operation. The morphological characteristics of rural settlements can be the combination of elements such as space, time, or time and space [15, 16]. The use of local map structure can reflect the internal relationship between the morphological characteristics of rural settlements from different angles. The partial diagram structure can be defined by the following formula:

$$Z_M = \frac{1}{G(x)} \sum_{M=1}^{n} F(x_M, y_M) H(y_M), \tag{8}$$

where M is the spatial position of the output image, and its value can be calculated by x_M and each corresponding point y_M of other images. x is each morphological feature of the input image, and y is the output image feature corresponding to x. The correlation function $G(x)$ can calculate the correlation between various image morphological features. The function H is mainly used for scaling the input image, and the function $G(x)$ is mainly used for normalizing the output morphological features.

The correlation function F can be expressed as follows:

$$F(x_M, y_M) = \sigma(x_M)^T \zeta(y_M), \tag{9}$$

where $\sigma(x_M)$ can be expressed as $\sigma(x_M) = Q_\sigma x_M$ and Q_ζ can be expressed as $\zeta(y_M) = Q_\zeta y_M$. The parameters Q_σ and Q_ζ can be corrected by convolution operation and training. The function H can be expressed in the following form:

$$H(y_M) = P_L y_M. \tag{10}$$

Function H is mainly used for linear transformation of the input feature H, and parameter P_L is processed by convolution operation and modified by training and learning. In order to facilitate each local image feature to be added to the neural network for processing, the output layer can be processed by residual operation, which can be expressed as follows:

$$B_M = P_A y_M + x_M, \tag{11}$$

where $P_A y_M + x_M$ represents a residual operation module. This residual processing method can input the local image feature into the pretrained network model and make the original model unaffected.

5.2. Result and Analysis. From the above analysis of the morphological characteristics of rural settlements, it is known that the morphological image of rural settlements has local characteristics, so it belongs to the local map structure. In order to explore the evolution process of the rural settlement form, the rural settlement form can be reconstructed by the graph neural network (GNN) algorithm. The local graph structure can use the local mean operation method to get the relationship between each layer of the neural network. Therefore, the morphological characteristics of rural settlements can be used as the input characteristic graph and processed by the local mean operation. The Comparison between settlement image processed by GNN and original image obtained by remote sensing is shown in Figure 6.

Because the morphological characteristics of rural settlements are mainly composed of space, time, or space-time and other elements, the use of this local map structure can reflect the internal relationship between the morphological characteristics of rural settlements from different angles. The statistical results of the evolution process of rural settlement morphology in 2020 are shown in Table 4. The evolution process of rural settlement morphology in 2020 is monitored based on the graph neural network algorithm, and the results are shown in Table 5. The monitoring results of various parameters in Table 5 are basically consistent with Table 4, which shows that the prediction model of the rural settlement form evolution process based on the graph neural network algorithm can better truly reflect the development and change process of the rural settlement form and scale.

From the above analysis results of the evolution process of the rural settlement form, it is known that the driving force for the formation and development of rural settlement mainly comes from external natural environmental factors,

Figure 6: Comparison between settlement image processed by GNN and original image. (a) Remote sensing image of rural settlement. (b) Settlement image map based on GNN.

Table 4: Actual statistical results of rural settlement scale in 2020.

Type	Area scope (hm^2)	Patch number	Patch proportion (%)	Patch area size (hm^2)	Total patch proportion (%)	Average patch area size (hm^2)
Alone	<0.2	26	0.57	2.34	0.01	0.09
Small	0.2–2	15	0.33	10.18	0.03	0.68
Medium	2–5	2248	49.26	6426.35	21.65	2.86
Large	>5	2275	49.85	23247.36	78.31	10.22

Table 5: Monitoring results of rural settlement scale in 2020 obtained by using this method.

Type	Area scope (hm^2)	Patch number	Patch proportion (%)	Patch area size (hm^2)	Total patch proportion (%)	Average patch area size (hm^2)
Alone	<0.2	25	0.55	2.25	0.01	0.09
Small	0.2–2	14	0.31	9.35	0.03	0.67
Medium	2–5	2235	49.10	6372.25	21.45	2.85
Large	>5	2278	50.04	23326.47	78.51	10.24

and it is regional location factors that promote the evolution of the rural settlement form, and it is economic and social development factors that play a leading role in the evolution process of the rural settlement form. There is a significant correlation and interaction between the external natural environmental factors represented by regional location and the economic and social development factors represented by farmers' per capita income. The natural environmental factors have a great influence on the evolution of the rural settlement form and last for a long time [6, 7]. Due to the urbanization factors, the production and life of rural settlements are affected, which accelerates the evolution of the form and scale of rural settlements. With the reorganization of rural population, land, social life, and industry, the scale and morphological structure of rural settlements will also change, which provides an internal driving force and conditional basis for the continuous change of rural settlements, which also promotes the dynamic and periodic evolution of rural settlements. Compared with 2010, the morphology of rural settlements changed significantly in 2020, among which the morphology of small settlements changed the most, and the number, proportion, area, and proportion of small patches decreased significantly compared with 2010. The number and area of medium-sized patches also decreased compared with 2010, but the average area of medium-sized patches increased, and the number and area of large patches increased compared with 2010. This shows that the form of rural settlements has changed greatly from 2010 to 2020, in which most small settlements have gradually changed to medium and large ones, and the form of rural settlements has changed significantly.

6. Conclusion

Due to the influence of various factors, the morphological structure and scale of rural settlements in mountainous areas were very complex. In order to track the evolution process of rural settlements in real time, an intelligent monitoring method of rural settlements morphological evolution process based on graph neural network algorithm was proposed in this paper. The traditional mathematical statistical method can only explore the evolution law of the rural settlement form from a qualitative point of view, but cannot quantitatively analyze the evolution process of the rural settlement form. Therefore, this paper used the graph neural network algorithm to extract and analyze the remote sensing

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9843539, 1 page
https://doi.org/10.1155/2023/9843539

Retraction

Retracted: Design of Pedagogy Course Information Sharing System Based on Wireless Sensor Network

Security and Communication Networks

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Jin, "Design of Pedagogy Course Information Sharing System Based on Wireless Sensor Network," *Security and Communication Networks*, vol. 2022, Article ID 7040441, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 7040441, 12 pages
https://doi.org/10.1155/2022/7040441

Research Article

Design of Pedagogy Course Information Sharing System Based on Wireless Sensor Network

Xianzhu Jin

College of Education, Jinggangshan University, Ji'an, 343009, China

Correspondence should be addressed to Xianzhu Jin; 9920190016@jgsu.edu.cn

Received 27 December 2021; Revised 18 January 2022; Accepted 31 January 2022; Published 1 March 2022

Academic Editor: Chin-Ling Chen

In order to improve the information sharing effect of pedagogy courses, it is necessary to improve the system information transmission network algorithm. This paper improves the wireless sensor network algorithm and analyzes and verifies the performance analysis model of the Aloha protocol in multihop underwater acoustic sensor networks (UASNs). Based on the improved performance analysis model of the Aloha protocol on the String topology network, the expected throughput of the String topology network and the average end-to-end delay can be obtained, and the transmission effect of pedagogy information in the sensor network can be improved. In addition, this paper constructs a pedagogy course information sharing system based on wireless sensor network to verify the information sharing effect of the model constructed in this paper. The experimental research results show that the system model proposed in this paper has certain effects.

1. Introduction

Education is a social activity system constructed artificially for the purpose of cultivating people, rather than a natural existence. Its core problem is to guide and regulate the development of people and to solve the problem of what kind of people to train and how to effectively train them. Pedagogy is a subject whose research object is education. Specifically, the purpose of pedagogy research on educational phenomena and educational problems is to reveal the laws of education. The focus of pedagogy research on educational values is often to discuss the multiple possibilities and value choices of the educational activity system, reality and ought, objective laws and the subjective and active creation of historical unity [1]. Pedagogy studies the art of teaching. The reason why teaching activities are called art is that teaching activities can and should be activities full of spirituality, emotion, and free creation. In view of the foregoing, we can think that the research object of pedagogy can have multiple choices, which can be the law of education or the value of education or the art of education. It can also be combined with the law of education, the concept of education value, the art of education, and its unity [2]. Therefore, pedagogy is a discipline that reveals the laws of education and explores educational values and educational art. Through the study of pedagogy, future teachers can master the laws of education, establish correct educational concepts, and master certain educational theoretical knowledge and scientific educational methods. This improves the consciousness and creativity of engaging in educational work, increases interest in educational work, deepens the love for education, and avoids blindness in work [3].

This article combines the wireless sensor network to construct a pedagogy course information sharing system, changes the traditional pedagogy course information processing method, and improves the effect of pedagogy teaching.

2. Related Work

Some experts pointed out that higher education currently cannot better interpret, predict, and guide actual teaching activities. In addition, it is a development trend that can be selected to gradually shift from problem exploration to theoretical creation. The key task should be to explore the laws of higher education, which should be regarded as a basic theoretical discipline [4]. Some studies proposes that in pedagogy, planning higher education as a subdiscipline has been unable to adapt to the current development needs of higher education theory, and higher education must be expanded into an independent first-level discipline. Higher education in the world is not regulated under the framework of pedagogy, and it has gradually tended to a separate discipline and has been continuously developing and growing [5]. Therefore, it is necessary to break the understanding category of traditional disciplines on the basis of the disputes on discipline theory and research field theory, and establish it as one of the contemporary disciplines [6]. The modern discipline paradigm of higher education can resolve and tolerate the contradictions and conflicts between the classical discipline paradigm and the research domain paradigm. The construction of higher education as a modern discipline and the realization of the redisciplinarity of higher education are the road to the transcendence of higher education [7]. Zawacki-Richter et al. [8] pointed out that when people think that higher education is not a discipline, they actually confuse the basis of discipline formation with the sound standards of disciplines. Higher education can adapt to the foundation and prerequisites for the formation of disciplines, but there are no sound indicators of disciplines, so it is still in the stage of continuous development. From this point of view, treating higher education as a developing discipline is a reasonable estimate based on the internal rules of the science and technology framework, and treating higher education as a developing discipline is based on the formation of the discipline. Scientific judgment is based on basic and sound indicators [9]. Lee pointed out that the current "discipline" and "collar city" continue to move closer, indicating that traditional discipline evaluation standards have become obsolete. To get out of the dispute between the "disciplines" and "fields" of higher education, we need to go beyond the traditional discipline evaluation standards and from the standpoint of specific disciplines, and from the height of the overall transformation of science, we should reunderstand the disciplines and their evaluation standards and insist that the discipline orientation is the reasonable and practical direction of higher education exploration [10]. Alonso proposes that higher education has basically completed the tasks of setting disciplines and discipline organization. The current task goal is to create a reasonable theoretical framework, that is, to form the internal mechanism of disciplines to promote or trigger the classification of disciplines. Then let it become an independent subject based on the normalization of theory [11]. Yang points out that higher education should become an interdisciplinary discipline system [12], and Ribeiro et al. believe that interdisciplinary discipline system has rich theoretical tension, which constitutes the consistency of higher education as a research category and discipline. From a standpoint, it not only lays a theoretical foundation for the discipline of higher education but also regulates the direction of its future development [13]. Rampton pointed out that multidisciplinary research methods have largely dismantled the traditional "three-independent" discipline standard's harsh treatment of higher education and resolved the identity crisis of higher education, while the "loose discipline" is precisely what. It is the multidisciplinary exploration that guides the future development path of the discipline of higher education [14]. Zhou puts higher education into the category of social disciplines [15].

Haejoong and Sangmin's study [16] is based on the fact that students study under pressure from further studies and employment. The state of "being in Cao Ying and sweating" and the teacher's helpless maintenance of classroom teaching in teaching only propose strategies to improve the teaching status of teachers and students. Lin et al.'s study [17] is based on understanding and improving the status quo of public pedagogy teaching methods in the context of new courses. Research on teaching methods and suggestions for improvement are carried out. Guan et al. [18] put forward a content-method comprehensive optimization model for the traditional teaching model that emphasizes the method level. Literature [19] aims to solve the practical problems of Public Pedagogy Textbooks and focuses on improving the practicability of Pedagogy Textbooks to improve students' interest, whether Pedagogy Textbooks can be practical, and how to achieve successful practicability.

Dilmurod and Fazliddin [20] put forward suggestions to improve the awareness of the importance of opening public pedagogy, rationally locate course objectives, and scientifically select and organize teaching content; Huang proposes to study the reform of high-level education courses from the connotation and current situation of teacher professional development [21]; Levchenko and Sadykova propose that according to the professional background of teachers, the new graduates cannot make good use of the pedagogical knowledge they have learned in practice and put forward suggestions for reforms in the nature, objectives, content, system, and class hours of the curriculum [22].

From the aforementioned analysis, it can be seen that the current research on pedagogy is still the traditional education model, and the intelligent Internet of Things teaching model English has not been incorporated into the pedagogy course information. Therefore, this paper designs the pedagogy course information sharing system based on the wireless sensor network.

3. Improved Performance Analysis Model of Aloha Protocol on String Topology Network

In order to improve the sharing effect of pedagogical teaching resources in the system, it is necessary to analyze the network topology to improve the system performance. First, a multihop Aloha protocol performance analysis model is established by using the String topology network.

The algorithm uses the String topology network to establish a multihop Aloha protocol performance analysis model and models the data flow arriving at each node as a Poisson process. We assume that λ_i is the arrival rate of data packets on node O_i. The probability P_i of the successful transmission of node O_i is the probability that the next hop node O_{i+1} successfully receives the data packet sent by O_i, and the expression is as follows:

$$P_i = \Pr\{O_{i+1}\text{Successfullyreceived}|O_i\text{Outgoingpackets}\}. \quad (1)$$

Based on the assumption that the interference radius of nodes in the String topology network is less than the distance between any two-hop neighbors, STNAM calls all nodes whose data packets can reach O_{i+1} as the competing node set of O_i, $C_i = \{O_iO_{i+1}O_{i+2}\}$, as shown in Figure 1. If the data packet of the node in C_i coincides with the data packet of O_i on O_{i+1}, the latter will not be able to reach O_{i+1} due to collision. Based on this analysis, the model further calculates the probability of each node to send out a package and derives a set of equations about P_i and λ_i, $i = 1, \ldots, n$.

However, there are two shortcomings in STNAM. First, STNAM assumes that the rate at which O_i sends out data packets is equal to λ_i. In fact, due to the limitation of the underwater acoustic modem half-duplex characteristics, when the underwater acoustic modem is sending data packets, other data packets arriving at the node will be discarded because they cannot be sent. Therefore, the rate of data packets sent by O_i is lower than that of λ_i. Second, for a data packet of O_i, not all data packets of nodes in C_i may collide with it on O_{i+1}. As O_i can only transmit one data packet at any time, there is no collision between O_i's data packets on C_i, that is, when a data packet of O_i arrives at O_{i+1}, the success of its reception has nothing to do with other O_i data packets. Therefore, STNAM amplifies the collision probability when the data packet is received. In fact, the algorithm only uses STNAM to evaluate the network performance and does not verify STNAM through simulation experiments, thus ignoring the aforementioned two points.

Before proceeding to the next step, the algorithm first distinguishes the data flow actually sent by each node from the data flow arriving at the node. We assume that the rate at which O_i actually sends data packets is $\lambda_{\text{transmit}}(i)$, and the probability that data packets can be sent on O_i is $P_{\text{transmit}}(i)$. If and only if O_i is currently in the sending state, O_i can send out data packets. Therefore, $P_{\text{transmit}}(i)$ is equal to the probability that no data packet will reach O_i within the data packet transmission time T, and it can be obtained:

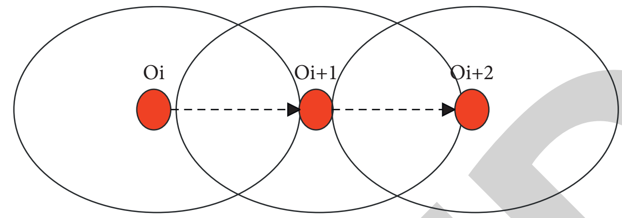


Figure 1: Schematic diagram of competing node set.

$$\begin{aligned}
P_{\text{transmit}}(i) &= \frac{e^{-\lambda_i T}\cdot(\lambda_i T)^0}{0!} = e^{-\lambda_i T}, \\
\lambda_{\text{transmit}}(i) &= \lambda_i \cdot P_{\text{transmit}}(i) = \lambda_i \cdot e^{-\lambda_i T}, i = 1, \ldots, n.
\end{aligned} \quad (2)$$

The data stream arriving at each node is composed of the data stream generated by the node itself and the data stream successfully transmitted by the previous hop node, and the rate at which each node generates data packets is λ. Therefore, the arrival rate of data packets on each node can be expressed as follows:

$$\begin{aligned}
\lambda_1 &= \lambda, \\
\lambda_i &= \lambda_{\text{transmit}}(i-1)\cdot P_{i-1} + \lambda, \quad i = 2, \ldots, n.
\end{aligned} \quad (3)$$

Incorporating formula (3) into formula (2), we can get the following equation:

$$\begin{aligned}
P_{\text{transmit}}(i) &= e^{-\lambda_1 T} \\
&= \begin{cases} e^{-\lambda T} & i = 1 \\ e^{-(\lambda_{\text{trummum}}(i-1)\cdot P_{i-1}+\lambda)T} & i = 2, \ldots, n \end{cases}.
\end{aligned} \quad (4)$$

From formula (4), the actual data packet rate of each node can be further expressed as follows:

$$\begin{aligned}
\lambda_{\text{transmit}}(i) &= \lambda_i \cdot P_{\text{transmit}}(i) \\
&= \begin{cases} \lambda e^{-\lambda T} & i = 1 \\ (\lambda_{\text{transmit}}(i-1)\cdot P_{i-1}+\lambda)\cdot e^{-(\lambda_{\text{munmum}}(i-1)P_{i-1}+\lambda)T} & i = 2, \ldots, n \end{cases}.
\end{aligned} \quad (5)$$

As the node does not consider whether the target receiving node is ready before sending, the vulnerability of Aloha protocol packet reception is twice the packet transmission time, that is, $2T$. In the String topology network, the success of the reception of O_i's data packet depends on the state of O_{i+1} when it reaches O_{i+1} and the arrival of other data packets during the reception of this data packet. If the downstream neighbor O_{i+2}'s data packet arrives at O_i during the fragile period of O_{i+1}'s data packet reception, the reception of O_i's data packet will fail due to collision. Moreover, if O_{i+1} initiates transmission during the fragile period of O_i's packet reception, O_i's packet reception will also fail due to blocking. If and only if O_i is idle during the vulnerable period of packet reception, the reception of O_{i+1}'s data packet can be successful. The aforementioned constraints are independent.

According to the assumption that the data packet arrives at the node and obeys the Poisson process, the probability that the node O_j does not send a data packet during the vulnerable period of the data packet reception, which is derived in formula (6):

$$\frac{e^{-\lambda_{\text{ramanit}}(j)\cdot(2T)}\cdot\left(\lambda_{\text{transmit}}(j)\cdot(2T)\right)^0}{0!}=e^{-\lambda_{\text{tracmitit}}(j)\cdot(2T)}. \quad (6)$$

Therefore, the probability of O_i data packet being successfully received can be expressed as follows:

$$\begin{aligned} P_i &= e^{-\left(\lambda_{\text{trucmili}}(i+1)+\lambda_{\text{rancuit}}(i+2)\right)\cdot(2T)}, \quad i=1,\ldots,n-2,\\ P_{n-1} &= e^{-\lambda_{\text{rasmit}}(n)\cdot(2T)},\\ P_n &= 1. \end{aligned} \quad (7)$$

In order to obtain the values of $\lambda_{\text{transmit}}(1)$, $\lambda_{\text{transmil}}(2),\ldots,\lambda_{\text{transmm}}(n)$, the n-dimensional nonlinear equations in formula (7) are used as constraints to solve the following objective function minimization problem.

$$\min_{\lambda_{\text{trasmit}}(1),\lambda_{\text{rasmit}}(2),\ldots,\lambda_{\text{trasmit}}(n)}\sum_{i=1}^{n}\left[F_i'(\Lambda)-\lambda_{\text{transmit}}(i)\right]^2. \quad (8)$$

Among them, $\Lambda=(\lambda_{\text{transmit}}(1),\lambda_{\text{transmit}}(2),\ldots,\lambda_{\text{transmit}}(n))$, and $F_i'(\Lambda)$ is calculated by iterating formulas (5) and (7).

The aforementioned minimization problem can be solved by the Nelder–Mead simplex method. We use MATLAB to solve the performance analysis model proposed in this section and give the analysis results.

Next, it is necessary to analyze the possible data loss in the process of data sharing to improve the effect of system resource sharing.

Reanalyze the data packet generated when the node modem sends it; the data packet will be discarded by the node because it cannot be sent. Although no collision actually occurs at this time, this situation can be regarded as a collision. We call this collision a "discard collision" and call the actual collision "ordinary collision." Unlike all data packets of O_i in a normal collision, the data packets of O_i that are being sent in the collision will not be lost. We assume that $P_{\text{success}}(i)$ is the probability that the arriving O_i data packet is successfully received by O_{i+1}, and the expression is as follows:

$$P_{\text{success}}(i)=\Pr\{O_{i+1}\text{Successfullyreceived}|\text{Packetarrival}O_i\}. \quad (9)$$

The other definitions of the model are the same as in Section 1, and the arrival rate of data packets on each node is as follows:

$$\begin{aligned} \lambda_1 &= \lambda,\\ \lambda_i &= \lambda_{i-1}\cdot P_{\text{success}}(i-1)+\lambda,\\ &= \lambda\cdot\left(1+\sum_{j=1}^{i-1}\prod_{k=j}^{i-1}P_{\text{success}}(k)\right), \quad i=2,\ldots,n. \end{aligned} \quad (10)$$

The algorithm reanalyzes the probability of successful transmission for each hop on the network. At this time, the success of O_i's data packet reception depends on the status of O_i and O_{i+1}. If and only if O_i does not currently send a data packet, the data packet that reaches O_i can be successfully transmitted, and no other data packet reaches O_{i+1} during the fragile period of the data packet reception. In other words, there are three situations in which packets arriving at O_i are lost. First, when the packet is generated, O_i is sending other packets. Second, when O_{i+1} is cross-talking receiving O_{i+2}'s data packet or O_{i+1} is sending data packet, the data packet arrives at O_{i+1}. Finally, when O_{i+1} is receiving the data packet, O_{i+1} starts to send the data packet or O_{i+2}'s data packet arrives at O_{i+1}. The aforementioned constraints are independent. Therefore, the probability of O_{i+1} being successfully received is as follows:

$$\begin{aligned} P_{\text{success}}(i) &= e^{-\left(\lambda_{i+1}\cdot P_{\text{trasmit}}(i+1)+\lambda_{i+2}\cdot P_{\text{trasmit}}(i+2)\right)\cdot(2T)-\lambda_i T}\\ &= e^{-\left(\lambda_{\text{transmit}}(i+1)+\lambda_{\text{transmit}}(i+2)\right)\cdot(2T)-\lambda_i T}\\ &= P_i\cdot P_{\text{transmit}}(i), \quad i=1,\ldots,n-2,\\ P_{\text{success}}(n-1) &= e^{-\left(\lambda_n\cdot P_{\text{trassuit}}(n)\right)\cdot(2T)-\lambda_{n-1}T}\\ &= e^{-\lambda_{\text{transmit}}(n)\cdot(2T)-\lambda_{n-1}T}\\ &= P_{n-1}\cdot P_{\text{transmit}}(n-1),\\ P_{\text{success}}(n) &= e^{-\lambda_n T}\\ &= 1\times P_{\text{transmit}}(n)\\ &= P_n\cdot P_{\text{transmit}}(n). \end{aligned} \quad (11)$$

Combining equations (10) and (11), we can get

$$\begin{aligned} \lambda_1 &= \lambda,\\ \lambda_i &= \lambda_{i-1}\cdot P_{\text{success}}(i-1)+\lambda\\ &= \lambda_{i-1}\cdot P_{i-1}\cdot P_{\text{transmit}}(i-1)+\lambda\\ &= \lambda_{\text{transmit}}(i-1)\cdot P_{i-1}+\lambda, \quad i=2,\ldots,n. \end{aligned} \quad (12)$$

We assume that U_i is the utilization rate on the O_i to O_{i+1} link. As the traffic in the String topology network finally converges to the gateway node through the last hop node, the throughput of the network directly depends on the throughput of the last link on the network. The utilization and effective throughput of the String topology network are denoted as $U(n)$ and $S(n)$, respectively, and the expressions are as follows:

$$\begin{aligned} U(n) &= U_n\\ &= \lambda_{\text{transmit}}(n)\cdot P_n,\\ S(n) &= \lambda_{\text{trausmit}}(n)\cdot P_n\cdot L\cdot\alpha. \end{aligned} \quad (13)$$

In the formula, L is the average length of the data packet, its unit is bits, and α is the average factor occupied by the data part of each data packet.

The end-to-end delay of a data packet in the network is equal to the sum of all transmission delays between the source node and the destination node plus the end-to-end propagation delay. We assume that the average end-to-end delay of the network is $\overline{\text{delay}}$, and the expression is as follows:

$$\overline{\text{delay}}=\sum_{i=1}^{n}\left[\left(\prod_{j=i}^{n}P_j\right)\cdot\left(\sum_{k=i}^{n}\left(T+\text{Distance}\frac{(i)}{c}\right)\right)\right]. \quad (14)$$

In the formula, c is the sound wave velocity in water, the unit is m/s, T is the data packet transmission time, the unit is s, and distance(i) is the distance between node O_i and O_{i+1}, and its unit is m.

4. Pedagogy Course Information Sharing System Based on Wireless Sensor Network

This article combines the third part of the wireless sensor network improvement algorithm and the actual needs of pedagogy course teaching to construct the pedagogy course information sharing system. Sensor real-time sharing is mainly divided into six main service modules (Figure 2):

Sensor classification service: When the sensor of the perception layer is connected to the middleware, the metadata information of the sensor information will be registered, including the sensor industry, protocol, corresponding topic in the message service, message sharing authority, location, and other information. Users can quickly find the corresponding sensor information according to their needs through the directory service. Data communication service: It collects sensor information in real time through a long connection with the sensor device terminal. Moreover, it parses the sensor data of different protocols, and then sends it to the Topic or Queue of the corresponding message queue service. Message queuing service: The message queuing service establishes a unique Topic and Queue for each sensor, and manages the messages of each sensor separately. At the same time, as a message transfer service container, it decouples data communication services and message push services, making the entire architecture highly scalable. Message push service: It monitors the sensor (monitors the topic corresponding to the sensor in the message queue service). When a message arrives, the message push service pushes the corresponding sensor message in real time according to the sensor information subscribed by the user, thereby realizing the real-time sharing of sensor messages. The user first uses the sensor classification service, locates the corresponding sensor according to the demand, and obtains the metadata information of the sensor. Then, the user establishes a connection with the message push service, sends the metadata information to the message push service, and the message push service establishes a subscription relationship based on the metadata information. Storage service: It is mainly used to persist sensor metadata description information and sensor historical information and other data to improve the reliability and stability of the entire middleware. At the same time, as a cache service, it enables some scenes with high real-time requirements to quickly read the corresponding information. Monitoring management service: It monitors the server status of the entire middleware to discover potential risks of system operation and improve the reliability of real-time sharing of sensor information.

As there are too many nodes in the wireless sensor network, if the nodes directly purchase spectrum from authorized users, it is too cumbersome and complicated. Therefore, the cluster head is generally used to trade with the spectrum sharing pool on behalf of the nodes in the cluster. Figure 3 shows the general model of spectrum sharing in wireless sensor networks, where the number of authorized users is not necessarily equal to the number of spectrum users in the spectrum sharing pool. In fact, because not all authorized users have free spectrum or not all authorized users are willing to share spectrum, K is generally less than or equal to M. In each round of transactions, the SP determines the spectrum price Pi and the sold bandwidth Di for each participating authorized user after obtaining the spectrum demand $\sum d$ of all cognitive users. Moreover, it notifies the corresponding spectrum trading information to the corresponding cognitive users.

In the one-master and multislave spectrum sharing model, as the bandwidth W that authorized users can provide is limited, and the spectrum demand of cognitive users changes, and the spectrum price P of authorized users generally changes with the relationship between spectrum supply and demand. Each cognitive user decides the bandwidth they purchase according to their own spectrum requirements and the spectrum price of authorized users, so as to maximize their own income. The spectrum sharing model at this time is shown in Figure 4:

When multiple authorized users sell spectrum usage rights to a cognitive user or a cognitive user system, the spectrum d that the cognitive user can purchase is limited. Authorized users can control their own spectrum prices to strive for cognitive users to purchase more bandwidth in order to obtain greater benefits. Ultimately, the bandwidth di that a cognitive user purchases from each authorized user depends on the spectrum price of the authorized user and the cognitive user's preference for spectrum. In the spectrum sharing model in Figure 5, the authorized user informs the spectrum management agency of the free spectrum with a bandwidth of Wi and its selling price Pi. The spectrum management agency then informs the cognitive users of the information of the M authorized users, and the cognitive users determine the bandwidth di they purchase from each authorized user according to their own spectrum requirements, spectrum preferences, and the spectrum price of the authorized users.

As shown in Figure 6, there are two competing relationships in multimaster and multislave spectrum sharing. One is that authorized users compete to sell spectrum to cognitive users. If the bandwidth provided by an authorized user is small or the frequency price is high, then cognitive users will be biased towards choosing other authorized users. Therefore, in the competition with other authorized users, each authorized user must consider its own spectrum price in order to obtain the maximum benefit. The second competition is that multiple cognitive users compete to buy the spectrum of authorized users. If many cognitive users choose to purchase spectrum provided by the same authorized user, the corresponding channel will become very crowded, which will lead to an increase in frequency prices. When the price of spectrum rises, some cognitive users with low competitiveness (mainly depending on the degree of preference of the cognitive users on the spectrum) will withdraw from the competition and choose other authorized

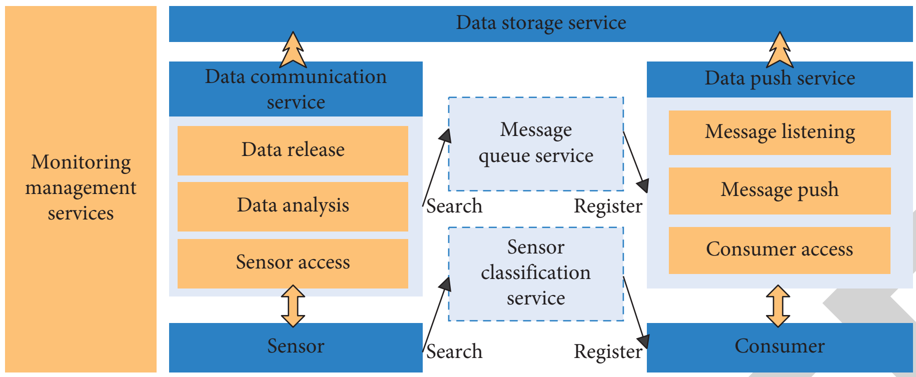


Figure 2: System function architecture diagram.

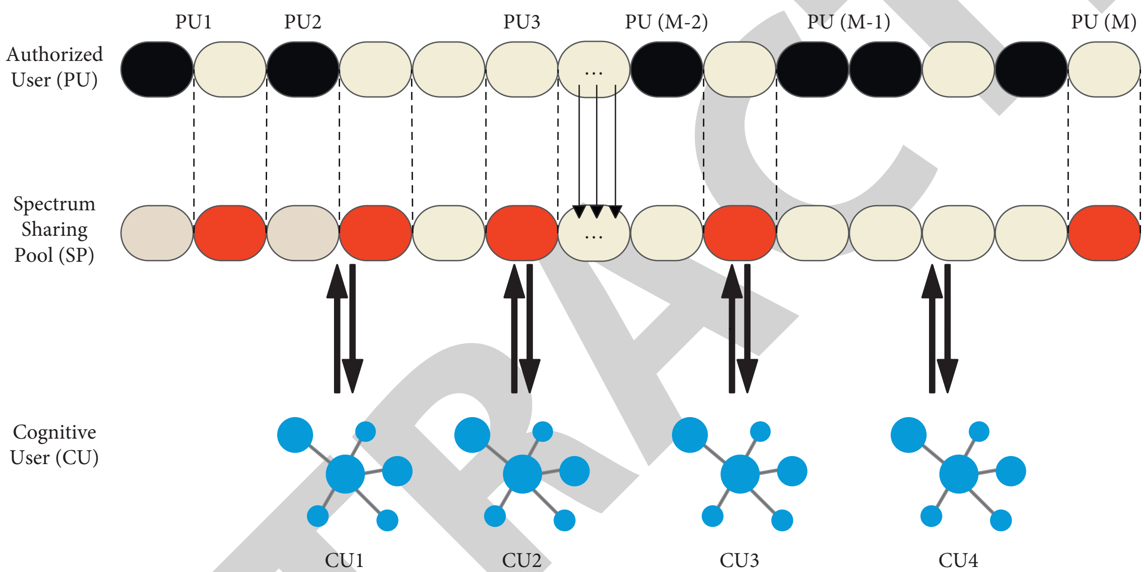


Figure 3: Spectrum sharing model in wireless sensor network.

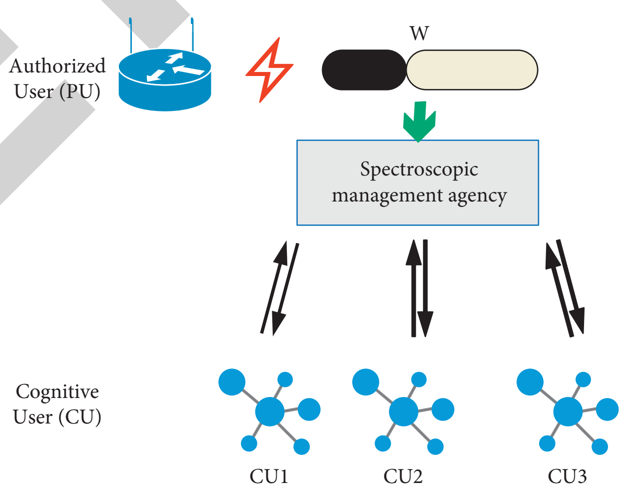


Figure 4: One-master and multiple-slave spectrum sharing model.

users with lower prices or relatively less competition. It is worth mentioning that the same authorized user in the model provides the same spectrum pricing for different cognitive users. Cognitive users can choose any authorized user, and which authorized user ultimately chooses depends on their own income.

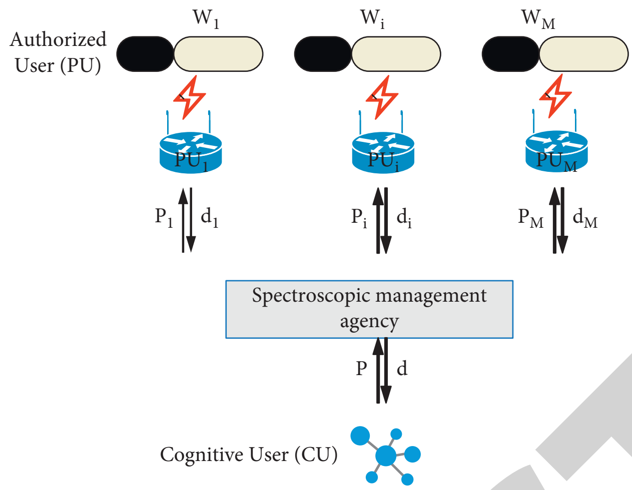


FIGURE 5: One-slave and multimaster spectrum sharing model.

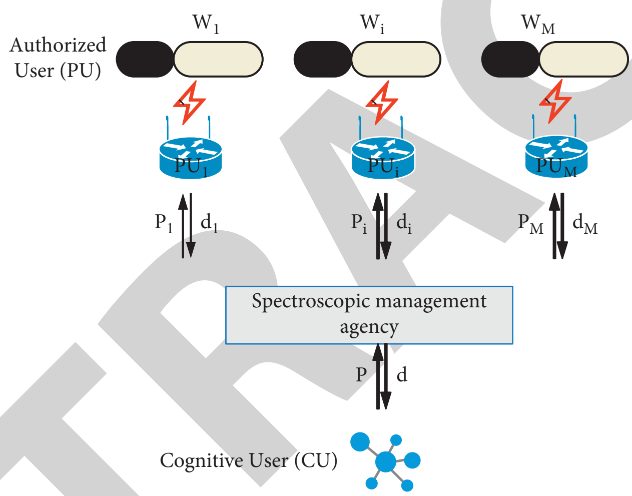


FIGURE 6: Multimaster and multislave spectrum sharing model.

At the level of pedagogy course information database, the traditional pedagogy course information data model contains all application data on a data server. However, with the development of pedagogy course information application, the amount of data is getting larger and larger, and the number of visits to a single library is getting larger and larger, so the database has to be split. This paper splits the database based on services, that is, different services use different databases. Originally, all data were concentrated in one database, but now it is split into multiple databases based on services, so that the pressure on one database is shared by multiple databases at the same time (Figure 7).

The pedagogy course information transmission channel specifically refers to the channel between the shared resource and the user of the resource in the sharing of educational digital resources. Media convergence is to closely connect different sharing channels to form a large transmission network, as shown in Figure 8. On this basis, the influence of multimedia integration on the sharing channels of community pedagogy course information digital resources is reflected in the expansion of the traditional resource transmission network and the innovation of the original human resources interaction channels.

The smooth implementation of education can not only be accomplished by content, but the construction of a virtual learning environment is equally important. An integrated platform must include a powerful backend that can support resource storage and transmission, application data calculation and analysis, and management and operation methods, and the content displayed in front of the audience is scientifically differentiated. As a resource management platform construction project, in addition to considering the management of various resources in the future, it is also important to consider the selection, cleaning, and integration of various heterogeneous resources precipitated by the school's history, especially the unified retrieval of various resources, personalized services, and the coconstruction and sharing of characteristic resources. An all-round resource sharing platform should adhere to the three-stage strategy of "improving the environment, expanding applications, and deep integration." Moreover, it uses "overall planning, unified standards, application-driven, and step-by-step

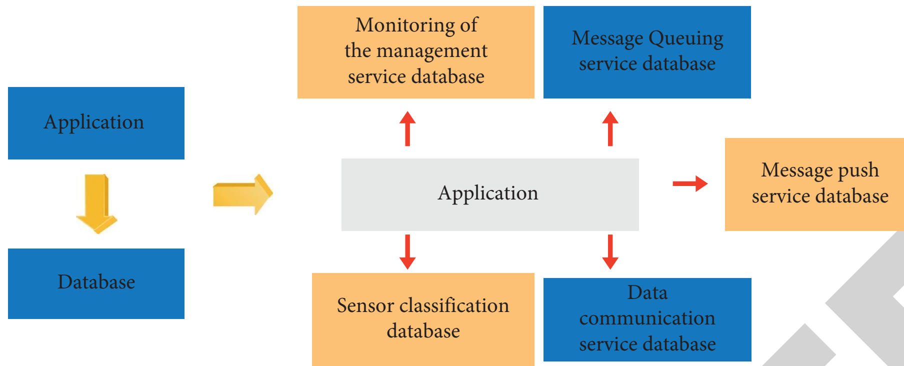


Figure 7: Subdatabase model of pedagogy course information database.

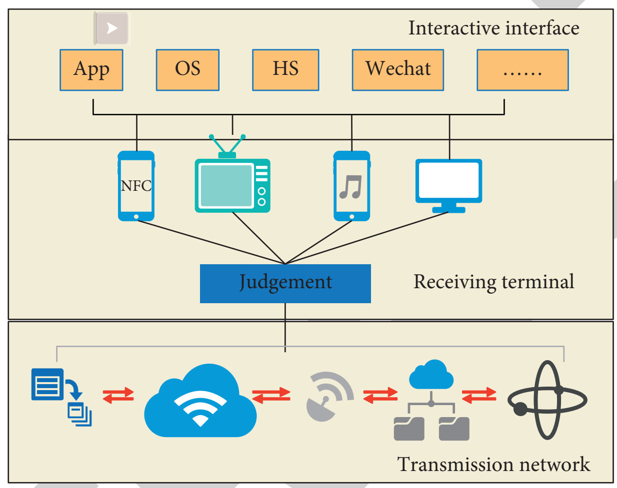


Figure 8: Fusion relationship between pedagogy course information transmission channels.

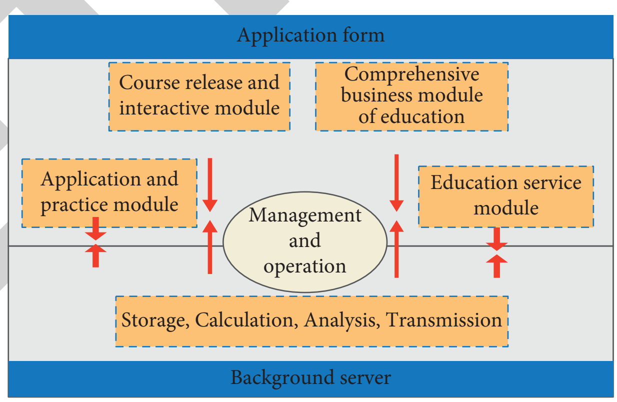


Figure 9: Structure relationship diagram of the content integration platform.

implementation" as its guiding ideology. Such a platform should consist of four parts, course release interactive module, community education integrated business module, application and practice module, and education service module. The structure relationship diagram of the comprehensive content integration platform is shown in Figure 9.

On the basis of the aforementioned research, the information sharing effect of the model constructed in this paper is verified, and experiments are designed to verify the model in this paper. This article collects a number of educational resources as experimental samples through a network library and builds a pedagogical information sharing system through a network platform. Moreover, this

Table 1: Information processing effect of wireless sensor network.

NUM	Information processing
1	97.0
2	96.1
3	94.5
4	96.8
5	95.5
6	96.5
7	93.5
8	91.8
9	93.0
10	96.7
11	92.0
12	90.7
13	94.0
14	93.0
15	94.9
16	94.5
17	94.6
18	92.9
19	93.4
20	92.5
21	90.5
22	94.2
23	93.0
24	96.3
25	90.5
26	95.2
27	93.2
28	94.4
29	94.9
30	91.7
31	95.3
32	96.2
33	95.3
34	94.0
35	93.4
36	92.3
37	96.2
38	95.7
39	93.6
40	95.5
41	96.3
42	91.2

Table 2: Teaching information sharing effect of course information sharing system.

NUM	Information sharing
1	87.0
2	91.0
3	87.3
4	89.9
5	91.2
6	89.0
7	90.9
8	87.2
9	86.0
10	90.2
11	87.9
12	86.3
13	85.3
14	86.8
15	85.9
16	90.2
17	89.3
18	91.7
19	87.6
20	91.9
21	90.6
22	85.7
23	86.0
24	84.6
25	91.9
26	90.8
27	88.8
28	92.1
29	90.3
30	89.1
31	88.8
32	91.8
33	91.2
34	88.8
35	90.9
36	85.8
37	85.4
38	87.1
39	86.9
40	87.0
41	86.7
42	84.4

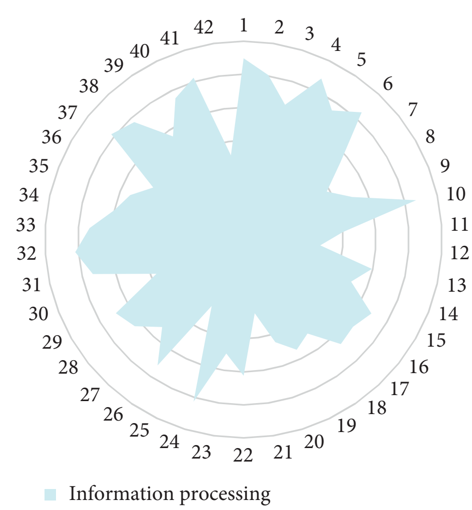


Figure 10: Statistical diagram of the information processing effect of wireless sensor networks.

paper verifies the information processing effect of the wireless sensor network and the teaching information sharing effect of the course information sharing system, and the results shown in Tables 1 and 2, Figures 10 and 11 are obtained.

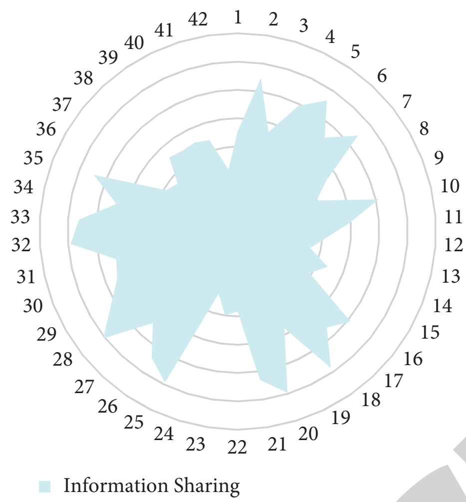


Figure 11: Statistical diagram of the teaching information sharing effect of the course information sharing system.

Table 3: Comparison of information sharing effects of pedagogy courses based on wireless sensor networks.

NUM	The method of this paper	The method of reference [12]
1	88.32	88.22
2	90.73	96.05
3	93.93	91.26
4	88.79	90.98
5	85.28	90.98
6	88.72	88.92
7	86.04	90.33
8	93.12	91.27
9	92.88	96.50
10	88.86	89.14
11	88.34	90.41
12	86.53	87.30
13	88.57	89.55
14	80.73	95.29
15	92.49	92.15
16	93.87	91.85
17	81.18	88.10
18	78.60	91.20
19	78.43	91.81
20	91.02	90.32
21	89.51	90.02
22	81.87	93.62
23	81.38	89.78
24	86.75	91.36
25	86.14	89.56
26	83.67	93.53
27	88.92	87.71
28	78.67	87.35
29	87.86	91.37
30	87.89	94.30
31	78.70	87.43
32	89.26	94.02
33	90.80	95.06
34	86.25	93.29
35	86.73	95.42
36	89.15	95.65
37	80.98	96.16
38	89.96	91.16
39	86.35	96.48
40	87.70	88.43
41	87.71	94.27
42	89.39	87.19

On the basis of the aforementioned research, the effect of the pedagogical course information sharing system based on wireless sensor network in pedagogical teaching information sharing is studied, and the research effect of this paper is compared with the literature [12], and the results are shown in Table 3 below.

Through the aforementioned experiments, it can be seen that the pedagogy course information sharing system based on the wireless sensor network proposed in this paper has a good course information sharing effect.

5. Conclusion

This paper first analyzes the current situation of pedagogy, summarizes the teaching problems, and draws the research purpose and research necessity of this paper. It summarizes and analyzes the research of experts and scholars, improves the wireless sensor network algorithm, analyzes and verifies many Jump the Aloha protocol performance analysis model in underwater acoustic sensor networks (UASNs), based on the improved String topology network. Aloha protocol performance analysis model can obtain String topology network throughput expectations and average end-to-end delay and improve the transmission effect of educational information in sensor networks, A wireless sensor network-based pedagogical course information sharing system is constructed, and the information sharing effect of the model constructed in this paper is verified.

In the face of fierce social competition, many teachers feel that they cannot adapt to the speed of knowledge renewal and the ever-changing teaching requirements. Therefore, their job satisfaction decreases, and they lack empathy and support for students, which in turn produces job burnout and stress. The most direct victims are students. The pedagogy course information resource is a key step in teacher preservice education. We must use a variety of effective methods to improve the sharing effect of pedagogy course teaching information, so that future teachers will enhance their confidence in education while learning knowledge and establish correct educational concepts. This article combines the wireless sensor network to construct a pedagogy course information sharing system, which changes the traditional pedagogy course information processing method and improves the teaching effect of pedagogy. The experimental verification shows that the pedagogy course information sharing system based on the wireless sensor network proposed in this paper has a good course information sharing effect.

Data Availability

The labeled data set used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This study was sponsored by Jinggangshan University.

References

[1] J. R. Mianroodi, N. H. Siboni, and D. Raabe, "Teaching solid mechanics to artificial intelligence—a fast solver for heterogeneous materials," *Npj Computational Materials*, vol. 7, no. 1, pp. 1–10, 2021.

[2] X. Li, "The construction of intelligent English teaching model based on artificial intelligence," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 12, no. 12, pp. 35–44, 2017.

[3] S. Zou, "Designing and practice of a college English teaching platform based on artificial intelligence," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 1, pp. 104–108, 2017.

[4] F. Kong, "Application of artificial intelligence in modern art teaching," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 13, pp. 238–251, 2020.

[5] M. Pantic, R. Zwitserloot, and R. J. Grootjans, "Teaching introductory artificial intelligence using a simple agent framework," *IEEE Transactions on Education*, vol. 48, no. 3, pp. 382–390, 2005.

[6] C. Yang, S. Huan, and Y. Yang, "A practical teaching mode for colleges supported by artificial intelligence," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 17, pp. 195–206, 2020.

[7] K. Kim and Y. Park, "A development and application of the teaching and learning model of artificial intelligence education for elementary students," *Journal of The Korean Association of Information Education*, vol. 21, no. 1, pp. 139–149, 2017.

[8] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education–where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–27, 2019.

[9] S. C. Yang, W. K. Vong, R. B. Sojitra, T. Folke, and P. Shafto, "Mitigating belief projection in explainable artificial intelligence via Bayesian teaching," *Scientific Reports*, vol. 11, no. 1, pp. 9903–9917, 2021.

[10] Y. Lee, "An analysis of the influence of block-type programming language-based artificial intelligence education on the learner's attitude in artificial intelligence," *Journal of the Korean Association of Information Education*, vol. 23, no. 2, pp. 189–196, 2019.

[11] J. M. Alonso, "Teaching explainable artificial intelligence to high school students," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 974–987, 2020.

[12] R. Yang, "Artificial intelligence-based strategies for improving the teaching effect of art major courses in colleges," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 22, pp. 146–160, 2020.

[13] P. Ribeiro, H. Simões, and M. Ferreira, "Teaching artificial intelligence and logic programming in a competitive environment," *Informatics in Education*, vol. 8, no. 1, pp. 85–100, 2009.

[14] V. Rampton, M. Mittelman, and J. Goldhahn, "Implications of artificial intelligence for medical education," *The Lancet Digital Health*, vol. 2, no. 3, pp. e111–e112, 2020.

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9894212, 1 page
https://doi.org/10.1155/2023/9894212

Retraction

Retracted: Deep-Learning-Based Motion Capture Technology in Film and Television Animation Production

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Wei, "Deep-Learning-Based Motion Capture Technology in Film and Television Animation Production," *Security and Communication Networks*, vol. 2022, Article ID 6040371, 9 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6040371, 9 pages
https://doi.org/10.1155/2022/6040371

WILEY | Hindawi

Research Article

Deep-Learning-Based Motion Capture Technology in Film and Television Animation Production

Yating Wei

Wuhan University of Communication, Wuhan 430205, Hubei, China

Correspondence should be addressed to Yating Wei; weiyating@whmc.edu.cn

Received 3 December 2021; Accepted 19 January 2022; Published 11 February 2022

Academic Editor: Chin-Ling Chen

With the popularity of King Kong, Pirates of the Caribbean 2, Avatar, and other films, the virtual characters in these films have become popular and well loved by audiences. The creation of these virtual characters is different from traditional 3D animation but is based on real character movements and expressions. An overview of several mainstream motion capture systems in the field of motion capture is presented, and the application of motion capture technology in film and animation is explained in detail. The current motion capture technology is mainly based on complex human markers and sensors, which are costly, while deep-learning-based human pose estimation is becoming a new option. However, most existing methods are based on a single person or picture estimation, and there are many challenges for video multiperson estimation. The experimental results show that a simple design of the human motion capture system is achieved.

1. Introduction

Motion capture technology is more mature and common in the film and television industry. After capturing the motion data of professional actors, doing specific processing, and then binding with the character model in the film and television works, we can get 3D virtual animation [1–3]. The currently used pose capture system is mainly divided into two categories: sensor capture and optical capture. The former is more mature, characterized by fast transmission speed and more accurate pose data; the disadvantage is the higher cost, and wearable devices are less convenient to use. In contrast, optical capture is the opposite, and there are two types of optical capture: unmarked and marked. The object of this paper is a markerless capture system, where a common 2D image or video is used as input to capture the human body's joint point data using target detection and feature extraction [4]. Although it is not yet widely used due to its unstable performance, its advantages such as ease of use, flexibility, and low cost should not be overlooked.

In recent years, numerous scholars at home and abroad have proposed considerable convolutional neural network models and other auxiliary methods for human pose estimation, covering single to multiperson, 2D to 3D, and picture to video [5]. However, human pose is a complex nonlinear model, and environmental noise, occlusion, and spatial depth ambiguity are the main hindrances to this task. If the input object is video data, it is also a difficult task to output a high frame rate and smooth and stable pose. Most of the existing methods are based on image-based 3D pose estimation [6], or for single-person videos [7]. The actual motion capture application objects are many times facing multiperson scenes, and the characters must have contact with the virtual physical space, so we propose a 3D multiperson estimation model from the video to meet the practical needs.

There are two general types of 3D pose estimation: one is regressive—regressing the 3D coordinates of the nodes directly from 2D data, which require the data to be 3D labeled, which is often difficult to obtain—and the other is the lifting type [8] where the two-dimensional pose is first obtained and then a mapping method is trained to lift it to the three-dimensional space on top of the two-dimensional one. The work in this paper will be centered on the lifting style, where current 2D estimation methods are relatively mature. And we will focus on the implementation of 3D estimation especially for multiperson targets. Time-domain convolution makes full use of key information at different time points of the video stream to infer 3D pose. Treating the human key point connection relationship as a graph structure is a prerequisite for implementing graph convolution, and then, the 3D relationship of bones can be extracted from global and local together. Previous approaches focus on feature learning and large-scale data training on pixel space, without making good use of a priori information such as human kinematics, spatial-physical relationships, and human topology [9], and for monocular 3D pose, inference relying on neural network learning alone is not sufficient. Some common problems, such as spatial relative position errors of multiple targets, contact with the ground with penetration and vacillation, unnatural tilt of pose and mutual occlusion, are found from the practical use of some models.

The remaining sections of this article are arranged as follows: Section 2 describes the modern mainstream motion capture system; Section 3 is the main content of this article is the design of film and television animation based on motion capture technology; Section 4 is the multiperson 3D estimation network; Section 5 is the experiment and evaluation; Section 6 is the conclusion.

2. Modern Mainstream Motion Capture System

The current mainstream motion capture systems can be divided into four categories: mechanical, electromagnetic, acoustic, and optical [10]. Optical-based motion capture systems mainly use multiple cameras to capture motion image sequences and trajectories and then accomplish the task of motion capture by identifying and tracking specific markers in the image information and using the motion information of these marker points to perform the 3D reconstruction.

2.1. Mechanical Motion Capture System. Mechanical motion capture system relies on mechanical devices to track and measure the motion trajectory. Mechanical motion capture systems generally consist of multiple joints and rigid connecting rods [1]. When the device is in motion, the position and trajectory of the rod end point in space can be derived from the angle change measured by the angle sensor and the length of the linkage.

2.2. Acoustic Motion Capture System. Acoustic motion capture system consists of a transmitter, a receiver, and a processing unit. The transmitter is a fixed ultrasonic generator, and the receiver consists of 3 ultrasonic probes arranged in a triangular pattern [11, 12]. This type of device is relatively low cost, but the capture of motion has a large delay and lag, real time is poor, the accuracy is generally not very high, and the sound source and the receiver cannot have large obscuring objects between, by noise and multiple reflections and other interference. Since the speed of sound waves in the air is related to air pressure, the corresponding compensation must also be made in the latter algorithm.

2.3. Optical Motion Capture System. Motion capture system is the most widely used and convenient system in the world. It uses multiple infrared cameras to capture objects from different angles. The software is then used to analyze the image coordinates of the marker points on the image, and the 3D reconstruction is performed using computer vision principles to derive the motion data of the marker points. The advantages of optical motion capture are a large range of performer activities, no cable, mechanical device limitations, easy to use, and high sampling rate [13].

3. Design of Motion Capture-Based Animation for Film and Television

Motion capture technology needs to build a skeleton model in order to describe the motion of the real human body, and all the skeleton models we store in the motion module database are shown in the figure below, and in order to let the captured motion data to drive the 3D human model, we need to combine the model with the captured motion data to achieve matching with the model, so as to drive the movement of the model. Finally, the model is matched with the captured data, and the model can follow the captured motion data to move, as shown in Figure 1.

In the past, the movie was to make the real into virtual, but nowadays, it is making the virtual scene into reality [14]. A film and television works using motion capture technology only need the following steps can quickly produce a film and animation works liked by the audience, and the actual operation steps are shown in Figure 2.

We use motion capture technology for film, television, and animation production, which can greatly improve the level of film, television, and animation production. It can greatly improve the efficiency of film and animation production, reduce the cost of film and animation production, and make the film and animation production process more intuitive and more vivid effects [11, 15–17].

4. Multiperson 3D Estimation Network

The current single-person 3D estimation and multiperson 2D estimation methods are relatively mature, while multiperson 3D estimation has many challenges. The study in [18] proposes a multiview approach, which has good results but requires specialized datasets and is difficult to collect data from everyday scenes. In this paper, we use monocular video data to achieve multiperson motion capture at a lower cost. Current time-domain convolution and graph convolution are two important methods to achieve pose estimation, and

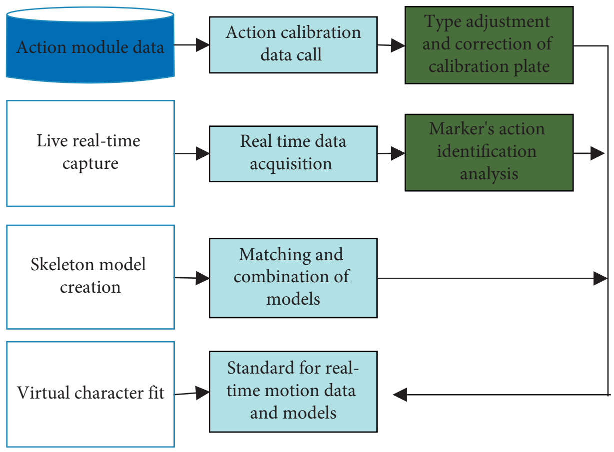


Figure 1: Motion capture-based film and television animation production process.

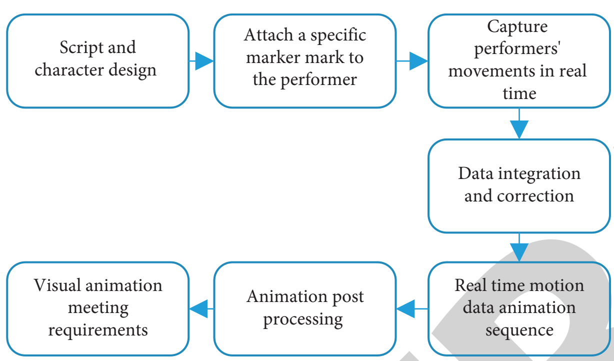


Figure 2: Motion capture technology for film and television production process.

combining the advantages of these two methods has also become a popular research direction [19]. Different targets in the real world have different distances from the camera, which is reflected in the image that the targets occupy different pixels, so the 3D distribution of multiple targets needs to be inferred from the 2D space, and the method used in this paper is depth estimation, and feeding the depth information to the time-domain graph convolution can achieve multiperson 3D estimation, which is the difference between multiperson and single-person.

4.1. Time-Domain Map Convolution. The time-domain convolution network proposed in [20] exploits the continuity of the post in the time domain and performs a void convolution in the time domain from the two-dimensional pose input to lift it to the three-dimensional space. Time-domain convolution is derived from RNN and LSTM. The study in [21] avoids the fact that RNN networks cannot be processed in parallel in the time domain, and the gradient length between the input and output is fixed so that the input sequences of different lengths can be trained stably without gradient vanishing and exploding. The network also employs null convolution to obtain long time-domain information without being limited to the top and bottom frames. The model works well for 3D estimation in the video but is not applicable to multiperson situations. Inspired from [22, 23], the application of graph convolution to time-domain convolution enables to obtain better spatial information of joint points. The form of skeletal joint point data is a graph structure, and the graph convolution methods before [24] basically learn the weight parameters of the convolution layer, while for the complex nonlinear transformation from two-dimensional key points to three-dimensional space, understanding the spatial relationship between joint points and the influence of each joint point on other joint points is another focus that cannot be ignored. Human skeletal dynamics is a nonregular and complex structure, and 3D pose estimation using graph convolution can overcome this limitation, and the model extracts both local and global pose information to accommodate occlusion situations. The local and global optimization is shown in Figure 3. Since the human pose in the video stream has continuity with both local dynamic and local static as well as temporary occlusion needs to rely on both the auxiliary of up and down frames and the graph convolution for pose space estimation, the graph convolution and time-domain convolution are performed simultaneously in the network structure proposed in this paper. To accommodate multiplayer estimation, corresponding improvements are made for time-domain convolution and graph convolution, respectively. Similar to the network structure in [25], four time-domain convolutional modules are constructed, each consisting of convolutional layers such as batch normalization, residual connectivity, ReLU loss function with random deactivation.

There is a graph convolution module between each two time-domain modules, considering the requirement of multiperson estimation, combining the depth information of the target on the basis of time-domain graph convolution, so that the multiperson pose has real spatial distribution and relative position relationship.

4.2. Depth Estimation. The study in [26] uses a top-down approach to first detect all the key points in the image and then combine the joints belonging to the same target according to the PAF (part affinity field) principle, which can be adapted to multihuman 2D pose estimation of video streams. Before the human pose estimation, there is another branching task-using Mask R-CNN [1] to first presample the person in the video and perform 2D pose estimation in the stationary extended state to obtain a standard skeletal model of each target for 3D pose estimation and correction unit. Then, all human targets in each frame of the video stream are detected, and the character markers with borders are output. The depth estimation of the targets is based on the 2D estimation of the joint point coordinates and the target detection with border regression. For each target segmented by the regression border, the pelvic key point P is selected as the human datum, and the target-to-camera distance (x_p, y_p, z_p) is calculated according to the imaging principle of the pinhole camera according to the method proposed in [11], where z_p denotes the distance to the camera. Lacking support conditions for estimating the distance only from

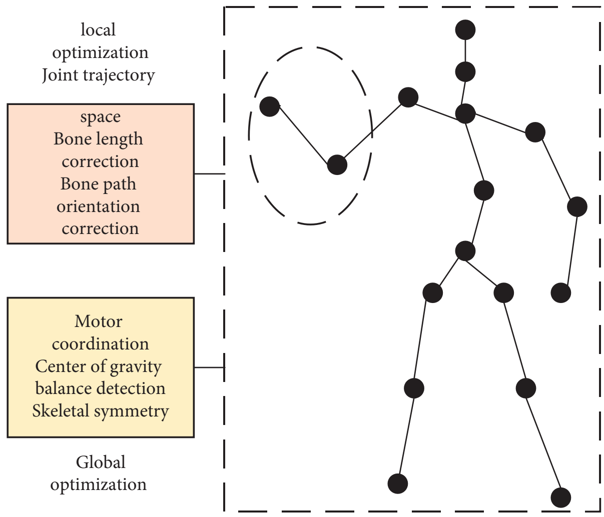


Figure 3: Schematic diagram of the structure of local and global optimization of the human posture.

two-dimensional images, a new measurement scale was designed k as follows:

$$k = \sqrt{\alpha_x \alpha_y \frac{A_{\text{real}}}{A_{\text{img}}}}, \quad (1)$$

where α_x, α_y denotes the focal length divided by the distance factor of X-axis and Y-axis, respectively, and A_{real} and A_{img} denote the area of the human body in real space (in units: mm^2) and image space (in units: pixel^2), respectively. For a given camera parameter, k approximates the absolute depth from the target to the camera by the ratio of the area of the real space to the area of the imaging space. The distance d can then be obtained according to the imaging principle as follows:

$$\begin{aligned} d &= \alpha_x \frac{l_{x,\text{real}}}{l_{x,\text{img}}} \\ &= \alpha_y \frac{l_{y,\text{real}}}{l_{y,\text{img}}}. \end{aligned} \quad (2)$$

Here, $l_{x,\text{real}}$, $l_{y,\text{real}}$, $l_{x,\text{img}}$, and $l_{y,\text{img}}$ denotes the length of the projection of the target on the coordinate axis in the actual space and the image space, respectively. However, this approach has drawbacks; for example, adults and children are in different depth cases, but the sizes are similar on the photos, and the depths output by the depth estimation module are the same. The solution is to involve the depth information in the training of the neural network, which in turn causes difficulties for the existing dataset and requires the addition of depth annotations, and eventually, a new dataset with a small portion of annotated data and a large portion of unannounced data is selected to train the network.

4.3. Contact Calibration. The study in [13] used a method of training neural networks to determine the foot-ground contact, although there is no dedicated public dataset currently available, for which they produced a new dataset. However, on the one hand, producing such a dataset requires a large investment, and on the other hand, the action relationship between the human body and the ground is still difficult to be fully described only by explicit image annotation, while dynamical analysis is an important reference for optimizing the contact problem, so it is proposed to optimize the contact relationship with the help of dynamical analysis in the absence of data and insufficient network training, which also helps to correct the human body imbalance in the next Step [14–16], and a simplification is done as shown in Figure 4. The human body model is simplified into eight parts such as head, torso, left arm, right arm, and left and right leg, and the whole body mass M is initialized to 70 kg, and the mass mi of each part is done according to the general proportion approximate distribution. The contact force between the human body and the ground only considers the internal forces generated by gravity and acceleration of motion. The support force of the ground on the human body comes from the biped, which is represented by $f_r, f_l \in \mathfrak{R}^3$, respectively, and the support force and the combined force of the human body are a pair of interacting forces. Human body dynamics are analyzed to find f_r and f_l.

$$\begin{cases} F = \sum_{i=0}^{J} \left(m_i \ddot{q}_i + m_i g \right), \\ \sum_{i=0}^{J} m_i \ddot{q}_i = \text{M}\ddot{\text{C}}, \\ \text{M}\ddot{\text{C}} = f_r + f_l + Mg, \end{cases} \quad (3)$$

where $q_i \in \mathfrak{R}^3$ are the coordinates of each part of the simplified model of the human body, obtained by averaging the coordinates of the corresponding joints; $\ddot{q}_i$ are their second-order derivatives; and $C, \ddot{C} \in \mathfrak{R}^3$ are the coordinates of the center of mass of the human body and their second-order derivatives.

The initial ground representation needs to be obtained along with the 2D pose estimation. Since it is difficult to label all the training data and, in addition, it is assumed that the plane in which the target is located is a flat area without undulations, only a portion of the data is labeled, which is used to obtain the initial ground on the one hand and to improve the adjustment of the correction unit to the contact state with the ground on the other hand.

Compared to the task of estimating the pose, the acquisition of the ground is relatively simpler and easier. A direct way to determine the foot-ground contact is to check the distance between the coordinates of the foot joint point and the initial ground d. And according to the kinetic analysis, the foot is in contact with the ground when the force of the ground on the body is greater than zero. If the output of contact markers estimated from 2D is wrong, or if the markers are correct, but the output 3D pose is penetrating or vacating with the ground, the pose space position needs to be corrected in combination with the kinetic

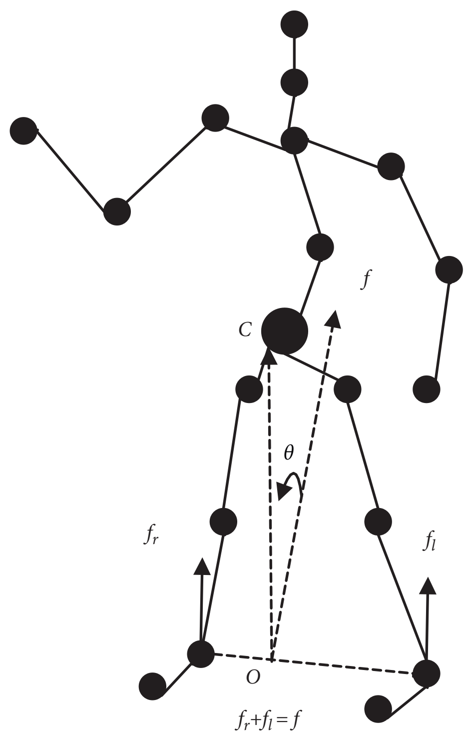


Figure 4: Human body mechanics analysis.

conditions. All joint point coordinates need to be flattened up or down by an offset distance Δd.

4.4. Balance Correction. For the human body balance state, there are static more dynamic. The static situation only needs to consider the effect of ground support and gravity, while the focus is on the dynamic balance. Each joint point of the human body generates momentum, acceleration, rotational moments, and complex forces between each other, and frictional forces on the ground in addition to pressure, which together constitute the equilibrium or imbalance of the human body. In order to maintain the postural balance, the study in [15] proposed controlling the human body balance by momentum calculation.

Motivated by these methods, a geometry-based balance control method is proposed in this paper. Combined with the prediction of the center of mass of the human body, the human body geometry is used to correct the human body balance state. First of all, it is still to determine whether the human body is out of balance, and the prerequisite of the judgment is that the current posture is in contact with the ground, disregarding the action of itself in the vacant state for the time being. The following model is used to analyze the conditions required for human equilibrium. The location of the center of mass C should be near the extension of the combined force f of the ground force f_r, f_l on the feet, and O is the intersection of f and the ground. The angle θ formed by the center of mass, O and f is within the threshold value, and the θ obtained from the imbalanced posture deviates significantly from this threshold value. Based on the measurements, this threshold is taken to be 5°.

$$\cos\theta = \frac{\overrightarrow{\mathrm{OC}}.\overrightarrow{f}}{|\overrightarrow{\mathrm{OC}}|.|\overrightarrow{f}|}. \tag{4}$$

Let the vector $\overrightarrow{k}\perp\overrightarrow{\mathrm{OC}}$ and $\overrightarrow{f}$, except for the joint points of the foot, and rotate around k by the angle θ.

$$\begin{aligned}\overrightarrow{k} &= (x, y, z),\\ \overrightarrow{\mathrm{OC}} &= (a, b, c),\\ \overrightarrow{f} &= (n, p, q).\end{aligned} \tag{5}$$

From the system of chi-squared equations,

$$\begin{cases} ax + by + cz = 0,\\ nx + py + qz = 0.\end{cases} \tag{6}$$

The vector k can be obtained as the vector from the original node to the point O is P, which is rotated around the unit vector u of k to obtain a new node, set as p':

$$p' = p\cos\theta + (u\cdot p)u(1-\cos\theta) + (u\times p)\sin\theta. \tag{7}$$

Loss function: let vector $\tilde{b}_k$ denote the standard bone length between the nth and mth joints, and b_k be the bone length output from the pose estimation network $b_k = \|p_n - p_m\|$. The loss of bone length prediction is as follows:

$$E_{\text{bone}} = \sum_{k=1}^{b}\left\|b_k - \tilde{b}_k\right\|. \tag{8}$$

The predicted loss of joint node coordinates is represented by L2 distance; P_{j-2D} and P_{j-3D} are the predicted 2D/3D joint point coordinates, and $\tilde{P}_{j-2D}$ and $\tilde{P}_{j-3D}$ are the real 2D/3D joint coordinates:

$$E_{\text{point}} = \frac{1}{J}\sum_{j=1}^{J}\left[\left\|P_{j-2D} - \tilde{P}_{j-2D}\right\| + \left\|P_{j-3D} - \tilde{P}_{j-3D}\right\|\right]. \tag{9}$$

5. Experimentation and Evaluation

5.1. Data Sets and Evaluation Mechanisms. The Human3.6 M dataset is the most commonly used dataset for human pose estimation tasks. It contains 3.6 million single-person video frames captured by a motion capture system in an indoor environment with 11 professional action actors showing 15 daily behaviors (e.g., walking, standing, talking) that can be adapted to single-person pose estimation and camera center coordinate prediction tasks. Based on previous experience, parts 1, 5, 6, 7, and 8 of the dataset are used for training and 9 and 11 are used for testing. The MuPoTS-3D dataset is a multiperson 3D pose estimation dataset that contains more than twenty indoor and outdoor scenes. The realistic 3D motion of each person in the video is derived from a multiview masterless capture that can be adapted to both human-centered and camera-centered coordinate systems. MuCo-3DHP is another dataset for multiperson 3D estimation, which is derived by combining the MPI-INF-3DHP

Table 1: Comparison with other methods using the Human3.6 M dataset under the MPJPE evaluation criteria.

Method	Dir	Dis	Eat	Phon	Pose	Pur	Sit	SitD	Smo	Phot	Wait	Walk
Dario	45.9	48.5	44.3	47.8	51.9	57.8	46.2	45.6	59.8	68.5	50.5	46.5
Cai	46.5	48.8	47.6	50.9	52.9	58.6	58.3	48.3	45.8	59.2	64.4	50.7
Moon	51.5	56.8	51.2	52.2	55.4	47.7	50.9	53.3	68.5	54.7	58.6	60.2
Xu	38.2	44.4	42.8	43.7	47.6	60.3	42.0	45.4	53.2	60.8	46.4	43.5
This article contains 5 frames	43.7	42.3	44.0	45.9	50.1	54.3	54.3	41.8	48.3	54.4	59.2	49.1
This article contains 9 frames	48.3	45.6	49.4	46.7	52.9	57.1	42.4	50.4	56.2	62.3	56.3	62.5

Table 2: Quantification of prediction accuracy on the MuPoTS-3D dataset.

Method	MP TPE	P-MPTPE
Basic network	55.4	47.5
Posture correction	51.3	43.7
Contact correction	52.1	43.6
Balance correction	52.4	42.9
Whole network	51.4	42.3

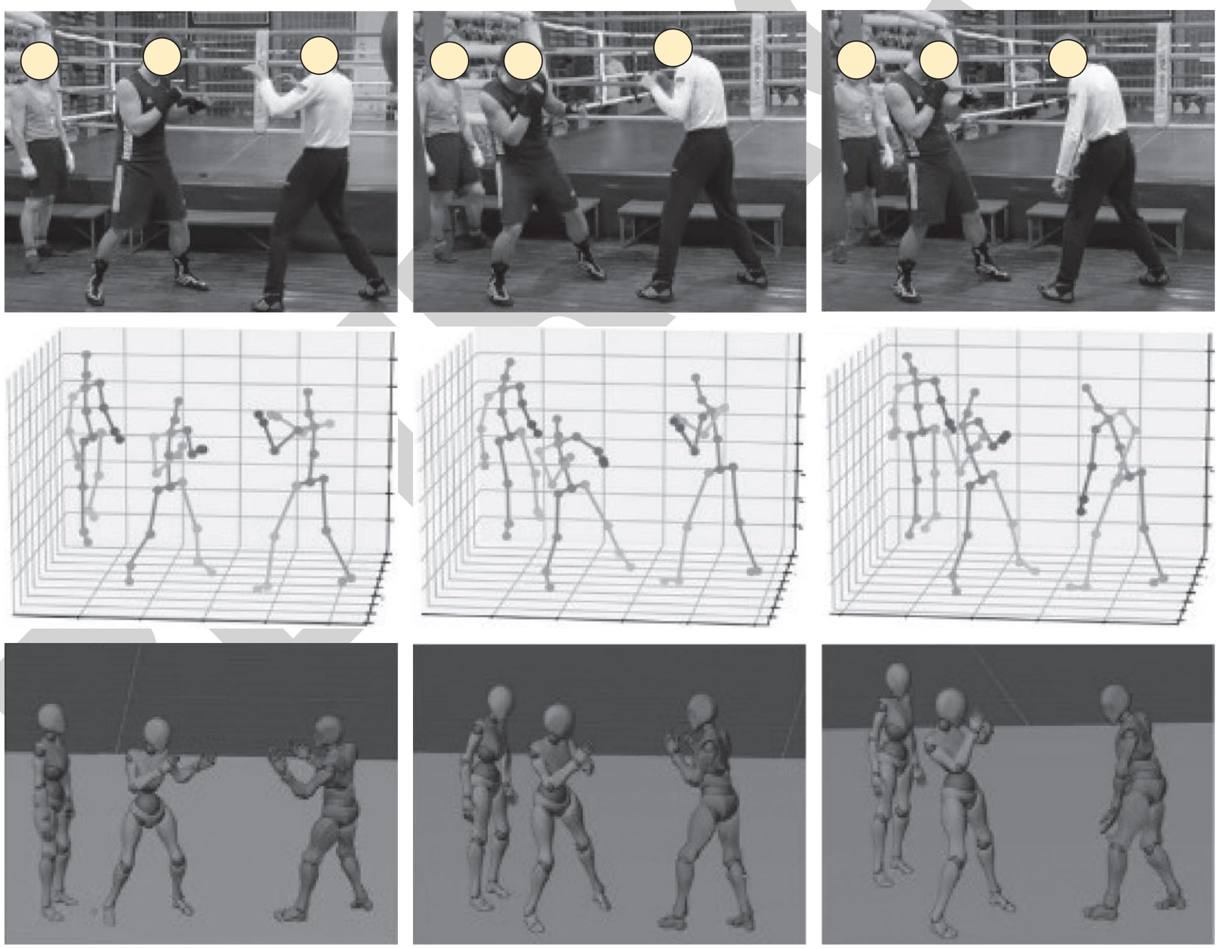


Figure 5: Motion capture visuals on the COCO2017 dataset.

3D single-person dataset. The MuPoTS-3D dataset is used for testing, and MuCo-3DHP is used for training.

Evaluation protocols: there are two evaluation protocols that are widely used: the first one calculates the mean error of predicted pose and true pose joint point coordinates (MPJPE), and the second one is the mean error after alignment [11] and this mechanism is called PA MPJPE [27].

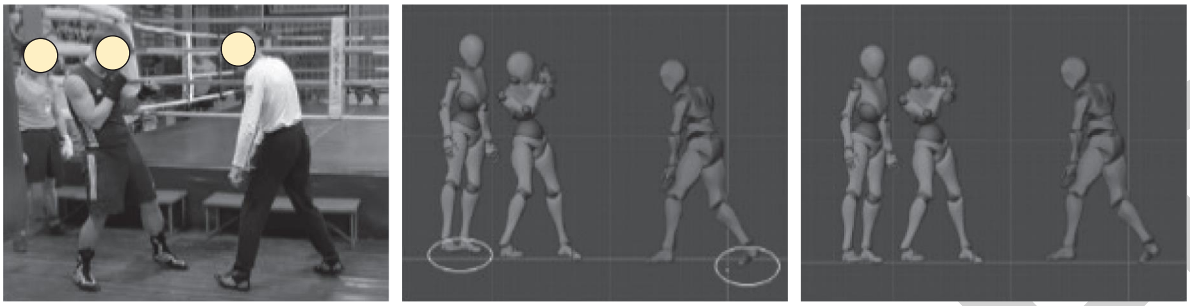


Figure 6: Contact correction effect.

Figure 7: Balance correction effect.

5.2. Realization Process. The Mask R-CNN exposed models used in this paper are pretrained on the coded dataset for target detection and age regression. For time-domain graph convolutional networks, their main parts are initialized with the publicly available ResNet-50, which is pretrained on the ImageNet dataset [15] is performed in training using a 256×256 size input image. The training of the pose estimation network is divided into two phases: the first phase uses the human 3.6 M dataset, and the 5th and 64th frames of each video are used for testing, using the MPII dataset in addition to the Human3.6 M dataset, with half of each of the two datasets in each training set. The MuCo-3DHP and MuPoTS-3D datasets were used in the second stage, and half of the data in each batch was COCO data in order to enhance the dataset. The experiments were conducted using four NVIDIA 1080 Ti GPUs for 20 rounds of training in the time-domain graph convolution network.

Experimental results on publicly available datasets show that the pose accuracy of the model output in this paper reaches the level of single-person video estimation, and the accuracy and smoothness of the output are related to the number of targets and frame rate. The data in Table 1 show that the accuracy of the monocular multiplayer video estimation model proposed in this paper is comparable to that of current single-person estimation models.

5.3. Calibration Module Assessment. Ablation results of the three correction units were compared on multiple evaluation protocols, using two evaluation protocols [28]. The data show that pose correction makes the largest contribution to accuracy improvement, with limited contributions from leveling correction and contact correction, but it is critical to the practical application of motion capture. The effect of each component on the final results is shown in Table 2, the multiplayer motion capture results are shown in Figure 5, and the correction units are shown in Figures 6 and 7.

6. Conclusions

The paper mainly discusses the application of modern motion capture technology in the production of film and television animation. With the continuous development and improvement of motion capture technology, motion capture technology will surely get more and more important applications in our daily lives. The emergence of optical motion capture systems has greatly reduced the cost of filming and animation production and has made film and animation pictures realistic. The motion capture model in this article can capture the actions of multiple people in a nonrestricted environment. After the human body pose is corrected, it basically conforms to the real human motion. It works well in three-dimensional virtual characters. It does not require professional shooting equipment and markings. In the case of other auxiliary equipment, materials can be obtained in daily life, and the simplified design of the motion capture system is basically realized. We will focus on the realization of three-dimensional estimation, especially the multiperson goal. Time-domain convolution makes full use of the key information at different time points of the video stream to infer 3D pose. Regarding the connection relationship of key points of the human body as a graph structure is a prerequisite for realizing graph convolution, and then, the 3D relationship of bones can be extracted from the global and local.

In the future, we need to pay attention to the preprocessing of data, especially the selection of data features, and the optimization of network models.

Data Availability

The dataset used in this paper is available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest regarding this work.

References

[1] Z. Manyu, "Application of performance motion capture technology in film and television performance animation," *Applied Mechanics and Materials*, vol. 347-350, pp. 2781–2784, 2013.

[2] T. Lauthelier and M. Neveu, "Facial animation by reverse morphing on a sequence of real images: application to film and video production," *Annales des Telecommunications*, vol. 55, no. 3/4, pp. 143–148, 2000.

[3] R. Zeng, "Research on the application of computer digital animation technology in film and television," *Journal of Physics: Conference Series*, vol. 1915, no. 3, Article ID 032047, 2021.

[4] S.-F. Xie, "Study on the creative development of China cartoon & animation - focusing on changes in a production environment with supporting the policy of the state administration of press, publication, radio, film and television of China," *Cartoon and Animation Studies*, vol. 35, pp. 209–226, 2014.

[5] B. Feng, W. Li, J. Shi, and Z. Li, "Framework study on the three-dimensional long-distance running sport training based on the markerless monocular videos," *Revista de la Facultad de Ingenieria*, vol. 32, no. 11, pp. 454–461, 2017.

[6] E. Akpinar, "The use of interactive computer animations based on POE as a presentation tool in primary science teaching[J]," *Journal of Science Education and Technology*, vol. 23, no. 4, pp. 527–537, 2014.

[7] W. Jian, "The analysis of the virtuality of film and television animation art," in *Proceedings of the 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)*, pp. 119–121, IEEE, Ottawa, ON, September 2014.

[8] C. Cao, Y. Tang, D. Huang, W. Gan, and C. Zhang, "IIBE: an improved identity-based encryption algorithm for wsn security," *Security and Communication Networks*, vol. 2021, Article ID 8527068, 8 pages, 2021.

[9] Y. Guo, B. Li, and K. Fan, "Support vector machine wavelet blind equalization algorithm based on improved genetic algorithm," *Advances in Intelligent and Soft Computing*, vol. 149, pp. 161–166, 2012.

[10] J. X. Huang, W. U. Wei-Long, and C. J. Long, "Study of moving object detection in video and its application based on OpenCV," *Computer Technology and Development*, vol. 135, no. 3, pp. 556–561, 2014.

[11] Y. Tong, W. Cao, Q. Sun, and D. Chen, "The use of deep learning and VR technology in film and television production from the perspective of audience psychology," *Frontiers in Psychology*, vol. 12, p. 501, 2021.

[12] C. Mo, K. Hu, S. Mei, Z. Chen, and Z. Wang, "Keyframe extraction from motion capture sequences with graph based deep reinforcement learning," in *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 5194–5202, Chengdu, China, 2021, October.

[13] D. Zhou, X. Feng, P. Yi et al., "3D human motion synthesis based on convolutional neural network," *IEEE Access*, vol. 7, Article ID 66335, 2019.

[14] J. Lin, J. Cui, G. Shi, and D. Liu, "CG animation creator: auto-rendering of motion stick figure based on conditional adversarial learning," in *Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 341–352, Springer, Xian, China, November 2019.

[15] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: a survey of deep learning-based methods," *Computer Vision and Image Understanding*, vol. 192, Article ID 102897, 2020.

[16] M. Y. Zhang, "Application of performance motion capture technology in film and television performance animation," in *Applied Mechanics and Materials*vol. 347-350, , pp. 2781–2784, Trans Tech Publications Ltd, 2013.

[17] R. Zeng, "Research on the application of computer digital animation technology in film and television," in *Journal of Physics: Conference Series*vol. 1915, no. 3, IOP Publishing, Article ID 032047, 2021.

[18] N. Du and C. Yu, "Research on special effects of film and television movies based on computer virtual production VR technology," in *Proceedings of the 2020 International Conference on Computers, Information Processing and Advanced Education*, pp. 115–120, Ottawa ON, Canada, October 2020.

[19] N. Du and C. Yu, "Application and research of VR virtual technology in film and television art," in *Proceedings of the 2020 International Conference on Computers, Information Processing and Advanced Education*, pp. 108–114, Ottawa ON, Canada, October 2020.

[20] X. I. E. Tao, C. Zhang, and Y. Xu, "Collaborative parameter update based on average variance reduction of historical gradients[J]," *Journal of Electronics and Information Technology*, vol. 43, no. 4, pp. 956–964, 2021.

[21] R. Ge, H.-Y. Chen, T.-C. Hsiao, Y.-T. Chang, and Z.-Y. Wu, "Virtual reality, augmented reality and mixed reality on the marketing of film and television creation industry," in *Proceedings of the International Conference on 5G for Future Wireless Networks*, pp. 464–468, Springer, Tianjin, P. R. China, August 2020.

[22] L. Wang, C. Zhang, Q. Chen et al., "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 160–167, IEEE, Wanzhou, China, November 2018.

[23] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, "Immune multipath reliable transmission with fault tolerance in wireless sensor networks," in *Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 513–517, Springer, Xi'an, China, October 2016.

[24] J. Zeng, X. He, Y. Hu, Y. Zhang, H. Yang, and S. Zhou, "Research status of data application based on optical motion capture technology," in *Proceedings of the 2021 2nd*

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9781089, 1 page
https://doi.org/10.1155/2023/9781089

Retraction

Retracted: Mental and Emotional Recognition of College Students Based on Brain Signal Features and Data Mining

Security and Communication Networks

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. An and Z. Yu, "Mental and Emotional Recognition of College Students Based on Brain Signal Features and Data Mining," *Security and Communication Networks*, vol. 2022, Article ID 4198353, 10 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4198353, 10 pages
https://doi.org/10.1155/2022/4198353

Research Article

Mental and Emotional Recognition of College Students Based on Brain Signal Features and Data Mining

Shaobo An[1] and Zhifen Yu[2]

[1]*School of Mathematics and Statistics, Central South University, HuNan, ChangSha 410083, China*
[2]*Middle School Affiliated to Central South University, HuNan, ChangSha 410005, China*

Correspondence should be addressed to Zhifen Yu; 205138@csu.edu.cn

Received 1 December 2021; Revised 20 December 2021; Accepted 17 January 2022; Published 8 February 2022

Academic Editor: Chin-Ling Chen

Nowadays, people pay more and more attention to the psychological situation of college students. Using data mining technology to model and analyze the collected psychological data of college students is a research hotspot in psychology and computer science. In addition, the essence of human emotional change is the higher nervous activity in the cerebral cortex. Electroencephalography (EEG) has become an important feature signal for emotion recognition because of its high time resolution and portability and practicality. Therefore, to solve the problem that the accuracy and generalization of the existing research models are not ideal, a method of college students' psychological emotion recognition based on EEG signal features and data mining is proposed. Firstly, a feature selection method based on sparse learning is used to find out a few features from the high-dimensional feature space that contribute greatly to the reconstruction of category information so as to quickly acquire a few key emotion-related features. Then, the entropy-weighted clustering algorithm is combined with sparse learning feature selection, and the local structure of heterogeneous data is divided. Experimental results show that, compared with traditional methods, the proposed method has stronger applicability and higher accuracy of five categories of emotions, which provides a valuable reference for the evaluation of depression and anxiety of college students based on brain signal characteristics.

1. Introduction

College students' mental health and happiness are not only vital to their own growth but also have an important contribution to society. There is evidence that the mental health of most college students is vulnerable at present. Many studies show that college students all over the world have a high rate of stress and depression. Teenagers' psychological problems are increasing gradually. More than 20% of young people have experienced psychological barriers at some point in their lives.

In recent years, with the development of medical imaging technology, emotion recognition based on central nervous system signals has become a research hotspot. The emotion recognition method based on the central nervous system refers to emotion recognition by analyzing the differences of signals sent by the brain in different emotional states [1–4]. At present, the recognition method based on the central nervous system mainly uses EEG. The emotion recognition method based on EEG signal has become one of the hot research topics in the field of emotion computing, which has attracted the attention of many researchers at home and abroad [5–8]. Especially in recent years, with the rapid development of machine learning technology, combining EEG signals with machine learning technology to identify and analyze emotions is becoming the mainstream research method in this research field.

At present, the feature extraction methods of emotional EEG signals can be mainly divided into three kinds [9, 10]: first, feature extraction methods based on the time domain, frequency domain, and time-frequency domain; second, the feature extraction method using potential technology; the third is to use the nonlinear dynamics of the feature extraction method. Regarding the selection of EEG signals, Jing et al. [11] pointed out that due to individual differences in emotions, different subjects may have different responses to different emotional stimulation materials. Therefore, data with large differences between samples should be excluded to enhance the accuracy and stability of model recognition. According to the research by Zheng and Lu [12], differences in EEG signals of different emotions in different bands are also relatively significant, and the band separation of EEG signals is conducive to improving the accuracy of emotion recognition. The purpose of using feature selection for emotion recognition is to find the representative and strong correlation features so that the selected features can be better studied and classified and improve the accuracy of emotion recognition.

Generally, there are two problems to be considered when constructing an emotion recognition model; one is classification accuracy, and the other is generalization ability. Especially for EEG signals, which have great differences between samples, we should pay more attention to these two problems [13–15]. At present, the algorithms selected by most related literature to build emotion recognition models are basically the support vector machine (SVM), decision tree, random forest, and neural network. For example, Mitsukura [16] segmented the EEG data of one minute with sliding windows and calculated 9 features of energy, energy ratio, and spectral entropy in alpha and beta bands on each segment. The features are input into the random forest classifier, and a good four classification accuracy of an emotional state is obtained.

Emotion recognition methods based on EEG feature signals and traditional machine learning technology [17–20] mentioned above can achieve relatively good results in general, but their shortcomings are also obvious, which are mainly reflected in two aspects: (1) most feature extraction methods of EEG signals are finely divided in all time periods and frequency domains, so the dimension of features will become very high, which makes rapid feature selection an urgent problem to be solved; (2) in many cases, the traditional emotion recognition model cannot fully reflect the hidden relationship between EEG signals and emotional states. That is to say, the more abstract, deeper, and more discriminating relationship between the EEG signal and emotional state cannot be obtained only by isomorphic data analysis, so it is difficult to make a breakthrough in classification accuracy.

Therefore, in order to solve the above two problems, this paper proposes a method of college students' psychological emotion recognition based on EEG signal features and data mining. The main work includes the following: (1) introducing the idea of sparse learning, a feature selection method of the EEG signal based on sparse learning is proposed, which can quickly, accurately, and effectively select emotion-related features from high-dimensional features and provides strong support for the subsequent construction of the emotion recognition model; (2) combining the entropy-weighted clustering algorithm with the features obtained after sparse learning, the classification model of depression and anxiety is established. By dividing the local structure of heterogeneous data, and marking and sorting the importance of features, the accuracy and generalization of clustering are provided. Finally, aiming at the positive and negative emotion classification task, the accuracy rate of 68.35% is achieved on the experimental dataset, which proves the effectiveness of the method proposed in this paper.

The rest of the paper is organized as follows. In Section 2, a processing mode of an input EEG signal is studied in detail, while Section 3 provides the detailed feature selection method. Section 4 provides the detailed emotional identification method. Section 5 provides detailed results and discussion. Finally, the paper is concluded in Section 6.

2. Artifact Removal of EEG Signal

As a physiological signal, EEG reflects the information of potential changes caused by complex nerve discharges in the brain [21]. EEG artifacts seriously affect the acquisition of signals related to neural activity. Fast and efficient removal of artifacts is the premise of the subsequent emotional state interpretation. We have adopted the EEG artifact removal method based on prior information [22, 23], and the process is shown in Figure 1.

Firstly, by adding a small amount of specific artifact prior information, the problem of mismatch of artifact prior information is overcome. Then, wavelet independent component analysis (WICA) is used to separate artifacts. As a combination of wavelet transform and independent component analysis, the WICA method has the advantages of multiresolution and multidimensional analysis. Its basic idea is to map signals to the wavelet domain and then carry out independent component analysis in the wavelet domain. Based on correlation discriminant analysis, the artifact components can be automatically identified and removed. Only a small amount of artifact prior information is used to effectively separate and remove the corresponding artifact components. In the range of 1 Hz to 50 Hz, the power spectrum energy is obtained with 10 Hz as the bandwidth.

3. Feature Selection Method of EEG Signal Based on Sparse Learning

3.1. Algorithm Principle. In the traditional filtering feature selection method, when the dimension of features becomes very high, the amount of calculation will become very time-consuming. In recent years, the sparse learning algorithm [24] has shown a good application prospect in the field of signal analysis. Through proper modeling, sparse learning can be well applied to feature selection. Then, by using the fast optimization method in sparse model solving, the operation speed of feature selection in high-dimensional feature space can be effectively improved.

The purpose of sparse learning is a sparse decomposition of signals, and a few optimal signals can be selected from the dictionary to reconstruct the original signals. After multiplying high-dimensional features by sparse weighting coefficients, the process of feature selection by sparse learning can be regarded as the process of approaching category labels. The final distribution of sparse weighting coefficients

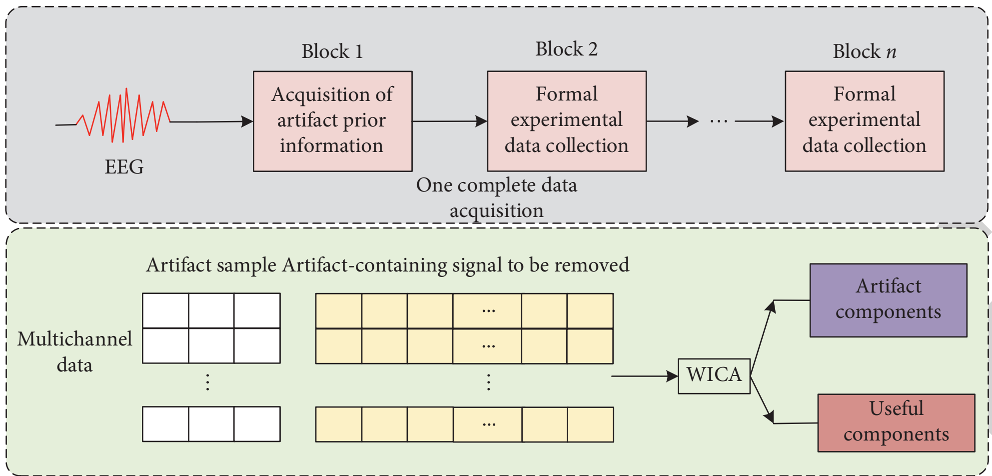


Figure 1: EEG artifact removal based on prior information.

shows the distribution of selected features. The overall method idea is shown in Figure 2.

The process of reconstructing category information by using high-dimensional feature space can be expressed by the following formula:

$$s_i = \mathbf{D}\alpha_i, \tag{1}$$

where s_i denotes a column vector related to class information, $\mathbf{D} = (v_1, v_2, \ldots, v_n)$ denotes a normalized feature vector group, and α_i denotes a sparse weighting coefficient. When the feature dimension is very high, $\mathbf{D} \in \mathbf{R}^{m\times n}$ can be regarded as an overcomplete dictionary matrix [25]. If the information of s is known, the problem of feature selection, i.e., solving $\alpha_i \in R^n$, becomes a problem of sparse learning.

Firstly, the category label matrix $\mathbf{S} = (s_1, s_2, \ldots, s_k)$ is constructed; k is the number of sample categories, where $s_i = (s_{i1}, s_{i2}, \ldots, s_{im})^T$. The specific expression is as follows:

$$s_{il} = \begin{cases} 1, & l-\text{th sample belongs to class } i, \\ 0, & l-\text{th sample does not belongs to class } i \end{cases}, \quad (l = 1, 2, \ldots, m) \tag{2}$$

Each column vector s_i in $\mathbf{S}$ can be sparsely represented by a feature matrix $\mathbf{D}$. And finally, a sparse matrix $\mathbf{A} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ is formed by solving the sparse coefficient α_i. The j-th row of the matrix $\mathbf{A}$ represents the contribution of the j-th feature to the reconstruction of the class information of each class of samples, and the sum of the absolute values of the j-th feature can be considered as the contribution to distinguishing the k-class samples. The feature importance score F is defined as follows:

$$F_j = \sum_{l=1}^{k} \left|\alpha_{jl}\right|. \tag{3}$$

As the importance score of the j-th feature for classification, the larger the F value, the greater the role of the corresponding feature for classification. Then, the features can be sorted according to the calculated F value, and the features with specified dimensions can be selected.

Sparse representation can be regarded as a problem of sparse decomposition of signals in an overcomplete dictionary. The basic idea is to select some atoms in the overcomplete dictionary (Φ) to represent the original signal y with optimal linear weighting. This is a process of sparse weighted approximation, which can be described by the following formula:

$$\hat{\mathbf{a}} = \operatorname{argmin}_{\mathbf{a}} \|\mathbf{\Phi a} - \mathbf{y}\|_2 + \lambda\|\mathbf{a}\|_0, \tag{4}$$

where λ is the regularization coefficient, which determines the degree of sparsity of the algorithm. $\hat{\mathbf{a}}$ is the required sparse representation coefficient.

3.2. Solution of Sparse Model. There are two kinds of commonly used sparse learning algorithms, namely, the relaxation algorithm and the greedy algorithm. The relaxation algorithm makes the problem gradually solvable by relaxing the constraints. The disadvantage of this method is that the complexity of the algorithm is high, and the complexity of the algorithm largely depends on the atomic library itself, which needs to be stored in the calculation process, so the requirement for storage space is relatively high. The greedy algorithm is to find out the supporting atoms of the signal to be reconstructed in turn through the idea of iteration. Its outstanding advantage is its fast operation speed, which has attracted extensive attention. Because the "backtracking" idea of the regularized orthogonal matching pursuit (ROMP) algorithm can better select the supporting atoms with global optimality, this paper uses this algorithm to solve the sparse model in feature selection.

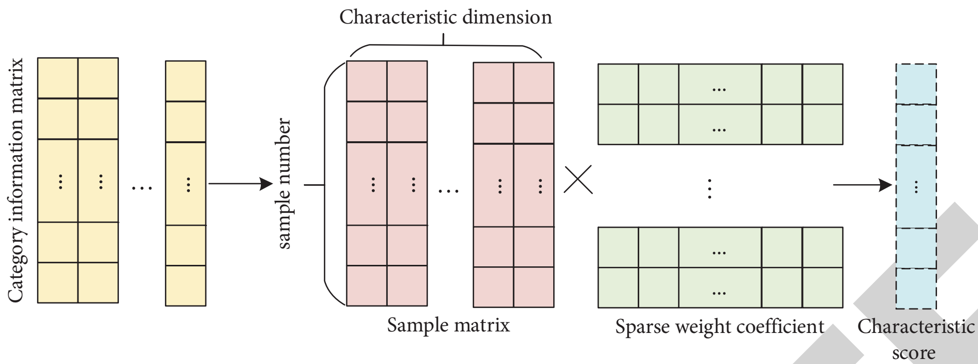


Figure 2: Feature selection based on sparse learning.

ROMP algorithm optimizes the selection mechanism of atoms. By adding the regularization step to realize the secondary screening of atoms, the global optimality of supporting atoms can be guaranteed to a great extent. The specific steps of the algorithm are as follows:

(1) The initial residual r is set as signal y. Firstly, $\Phi^T \mathbf{r}$ is calculated, and the largest n nonzero atoms are selected in $\Phi^T \mathbf{r}$ to form a set $\mathbf{J}$.

(2) In the set $\mathbf{J}$, the atoms are grouped according to the inner product $\mathbf{u}$ of atoms and residuals so that any two inner products u_i and u_j in each group satisfy $\|u_i\| \leq 2\|u_j\|$. The set with the largest energy is selected as $\mathbf{J}_0$.

(3) the atoms in $\mathbf{J}_0$ are added into the supporting atom set $\mathbf{X}$, the decomposition coefficient $\hat{\mathbf{a}}$ is calculated by using the least square method on the supporting atom set, and the residual error $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\alpha}$ is updated. Returning to the first step, the residuals are further decomposed until either the residuals are negligible or the number of supporting atom sets reaches a predetermined number.

4. Emotional Identification Method Based on Entropy Weighted Clustering

The research steps of emotion recognition based on EEG mainly include emotion induction, EEG signal acquisition, EEG signal pretreatment, feature extraction and selection, and emotion pattern learning and classification. Each step is essential for the research of EEG-based emotion recognition methods. The process of emotion recognition based on EEG is shown in Figure 3.

4.1. Representation of Sparse Learning Features. In order to reduce the spatial dimension of the EEG feature data obtained by solving the sparse model while retaining the features that are helpful for clustering, we use the sparse fractional feature representation. First, the input dataset is transformed with the objective of minimizing the L1 norm, assuming the set $\{x_i\}_{i=1}^n$, $x_i \in R^d$. Let $X = [x_1, x_2, \ldots x_n] \in R^{d\times n}$ and the sparse fractional coefficient for x be solved, as shown in the following formula:

$$\begin{gathered}\min_{S_i} \|S_i\|_1,\\ \text{s.t. } x_i = X' s_i,\end{gathered} \tag{5}$$

where X' is the matrix without column i.

The objective function of the sparse fractional coefficient is

$$S(r) = \frac{\sum_{i=1}^n \left(x_{ir} - (Xs_i)_r\right)^2}{\operatorname{Var}(X(r))}, \tag{6}$$

where $\operatorname{Var}(X(r))$ represents the accumulated difference of the r dimension feature.

Finally, the feature with the minimum objective function value $S(r)$ is selected as the clustering feature representation in this paper so that the unnecessary clustering features are removed.

4.2. Entropy Weighted Clustering Algorithm. Data mining refers to the process of revealing the implicit and potential value information from a large number of fuzzy, incomplete, and noisy complex data [26]. Among them, clustering analysis is a widely used data mining method at present, which can be regarded as a process of dividing data object sets. Recently, Fathian and Jafarianmoghaddam [27] proposed an optimization scheme for ad hoc networks based on the entropy-weighted clustering algorithm. Entropy is a measure of uncertainty that arises from the measurement of "disorder" in a physical thermodynamic system. After selecting appropriate features as the representation of clustering features, the clustering analysis process is started. For the construction of the emotional tendency recognition model, the main purpose is to calculate the clustering center matrix so as to get the membership of EEG signals to different emotional categories.

Therefore, if the object subjected to entropy-weighted clustering is set as $X = \{x_1, x_2, \ldots, x_N\} \subset R^D$, and the degree of membership expression is set as $u_{ij}^{(1)}$, the clustering center matrix is expressed as

$$v_{ik} = \frac{\sum_{j=1}^n u_{ij}^2 x_{jk}}{\sum_{j=1}^n u_{ij}^2}. \tag{7}$$

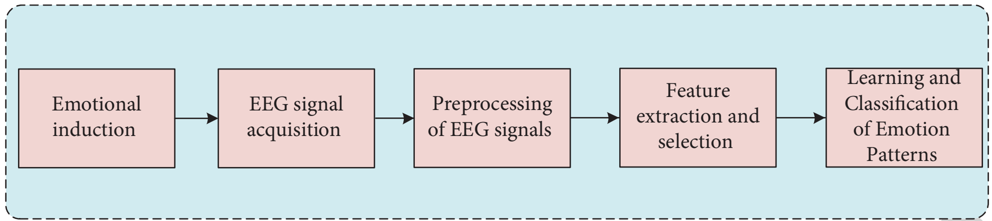


FIGURE 3: Process of emotion recognition based on EEG.

The datasets $U = [u_1, u_2, \ldots u_n]$ and $V = [v_1, v_2, \ldots v_n]$ are given, and the calculation of w_{ik} is set as

$$w_{ik} = \frac{\exp\left(-\sum_{j=1}^{n} u_{ij}^2 \left(x_{jk} - v_{ik}\right)^2 / \gamma\right)}{\sum_{s=1}^{d} \exp\left(-\sum_{j=1}^{n} u_{ij}^2 \left(x_{js} - v_{is}\right)^2 / \gamma\right)}. \quad (8)$$

Combining the above formula with the objective clustering function (6), the following formula is obtained:

$$\psi\left(w_{ik}\right) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^2 \sum_{k=1}^{d} w_{ik}\left(x_{jk} - v_{ik}\right)^2 + \gamma \sum_{i=1}^{c}\sum_{j=1}^{n} w_{ik} \log w_{ik} - \sum_{i=1}^{c} \lambda_i^w \left(\sum_{k=1}^{d} w_{ik} - 1\right). \quad (9)$$

The partial derivatives of formula (9) are calculated for w_{ik} and λ_i^w, respectively, and the result is made equal to 0; then,

$$\frac{\partial \psi\left(w_{ik}\right)}{\partial w_{ik}} = \sum_{j=1}^{n} u_{ij}^2 \left(x_{jk} - v_{ik}\right)^2 + \gamma\left(\log w_{ik} + 1\right) - \lambda_i^w = 0. \quad (10)$$

$$\frac{\partial \psi\left(w_{ik}\right)}{\partial \lambda_i^w} = \sum_{k=1}^{d} w_{ik} - 1 = 0. \quad (11)$$

Combining formula (10) and formula (11), we can get

$$w_{ik} = \frac{\exp\left(-\sum_{j=1}^{n} u_{ij}^2 \left(x_{jk} - v_{ik}\right)^2 / \gamma\right)}{\sum_{s=1}^{d} \exp\left(-\sum_{j=1}^{n} u_{ij}^2 \left(x_{js} - v_{is}\right)^2 / \gamma\right)}. \quad (12)$$

Therefore, the above formula (12) is a necessary condition of formula (5). Then, the entropy-weighted membership representation calculation method is defined as follows:

$$u_{ij} = u_{ij}' + u_{ij}'', \quad (13)$$

$$u_{ij}' = \frac{1}{\sum_{s=1}^{C}\left[\left(\sum_{k=1}^{d} w_{ik}\left(x_{jk} - v_{ij}\right)^2\right) / \left(\sum_{k=1}^{d} w_{sk}\left(x_{jk} - v_{sk}\right)^2\right)\right]}, \quad (14)$$

$$u_{ij}'' = \alpha \frac{N_i - N_j}{\sum_{k=1}^{d} w_{ik}\left(x_{jk} - v_{ik}\right)^2}. \quad (15)$$

In the above formula, N_j is calculated as follows:

$$N_j = \alpha \frac{\sum_{s=1}^{C}\left[1 / \left(\sum_{k=1}^{d} w_{sk}\left(x_{jk} - v_{sk}\right)^2\right) N_s\right]}{\sum_{s=1}^{C}\left[1 / \sum_{k=1}^{d} w_{sk}\left(x_{jk} - v_{sk}\right)^2\right]}. \quad (16)$$

5. Experiment and Result Analysis

5.1. Data Collection. A total of 80 college students aged from 18 to 25 were recruited in this experiment. Different emotions were induced by external stimuli in the form of pictures. Since the extraction of EEG signals is generally achieved by placing electrodes on the scalp, the signals are very weak, so hardware amplification by an amplifier is required after the EEG signals are extracted. The experimental data acquisition used GTEC's 16-channel G-USBamp system. Table 1 shows the specific settings of the acquisition parameters, and Figure 4 shows the placement locations of 16-channel EEG electrodes.

The letters in Figure 4 indicate where the electrodes are located. F represents the frontal lobe, C represents the center, T represents the temporal lobe, P represents the parietal lobe, and O represents the occipital lobe.

5.2. Emotional Data Analysis. For holistic analysis of the collected data, a depression scoring baseline was set. Those with the score below 53 points are defined as no depression, with the score between 53 and 62 as mild depression, the score between 62 and 72 as moderate depression, and the score above 72 as severe depression. The pie chart of the depressive mood data is shown in Figure 5. Health accounted for 70%, mild depression accounted for 16%, moderate depression accounted for 9%, and severe depression accounted for 5%. Generally speaking, most students are healthy. Similarly, anxiety was scored on a 50-point basis. The pie chart of depression data is shown in Figure 6. The EEG correlation between positive and negative emotions (depression and anxiety) is shown in Figure 7.

5.3. Data Mining Performance Verification of Entropy-Weighted Clustering. The main hardware configuration of the computer is 2.6GHZ CPU, 4G memory, and 500G hard disk. In order to save the initialization time cost of the algorithm, the dataset is firstly normalized, and the initial entropy-weighting coefficient is set.

First, the number of datasets is set to 15; that is, the initial target clustering center is 15, and the initial two-dimensional view is shown in Figure 8.

Table 1: EEG signal acquisition parameter settings.

Parameter name	Parameter value
Amplifier	16-channel G-USBamp system
Sampling frequency	512 Hz
High pass filter cutoff frequency	0.1 Hz
Low-pass filter cutoff frequency	60 Hz
Trap/notch frequency	50 Hz
Electrode position distribution	16 locations in 10–20 international standard system (as shown in Figure 4)
Grounding electrode position	Forehead
Reference electrode position	Right earlobe
Electrode material	Ag/AgCl
Recording software	g.Recorder

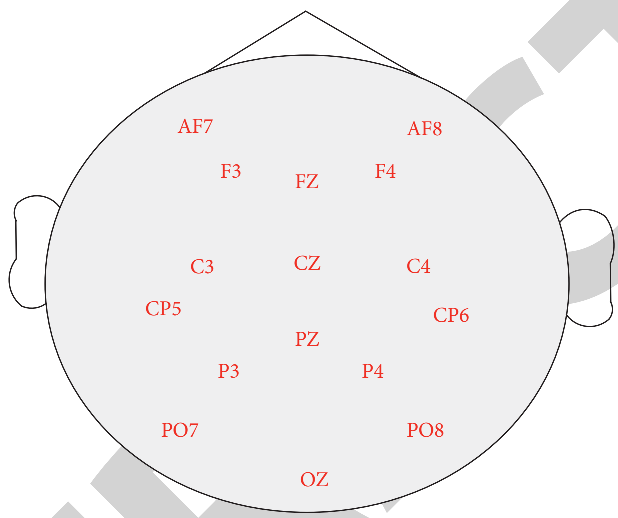


Figure 4: Schematic diagram of electrode position distribution.

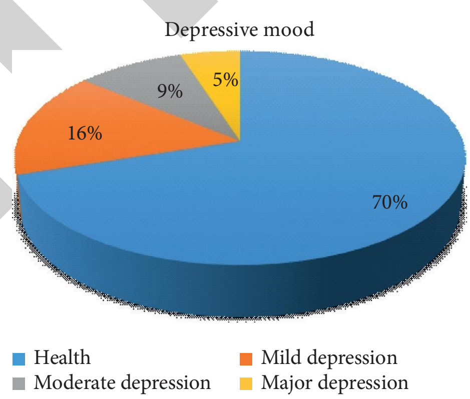


Figure 5: Depressed emotion pie chart.

After entropy-weighted clustering, the number of target clustering centers is reduced to three. In order to verify the convergence rate of the algorithm, the number of iterations of the algorithm was set to 2, 3, 9, and 16 times, respectively. The clustering results after 9 iterations are shown in Figure 9.

When the number of iterations was set to 2, 3, 9, and 16, the clustering centers became 9, 8, 6, and 4, respectively. The more iterations, the closer the number of target clustering centers was to 3. The number of iterations has a nonlinear relationship with the number of clustering centers. In order to reach the target clustering center number 3, the iteration is continued. The experimental results show that when the number of iterations is 17, the number of clustering centers becomes three, as shown in Figure 10. It can be seen that the number of clustering centers converges stably, and the clustering centers no longer change with the increase of iteration times, which verifies the generalization of the clustering algorithm.

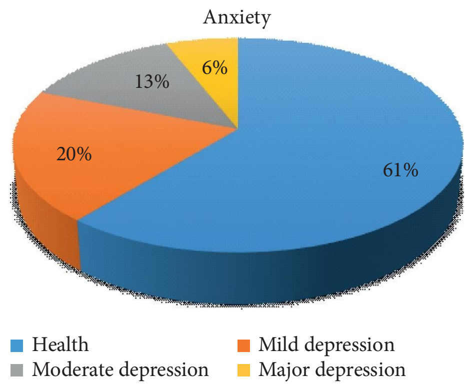


FIGURE 6: Anxiety emotion pie chart.

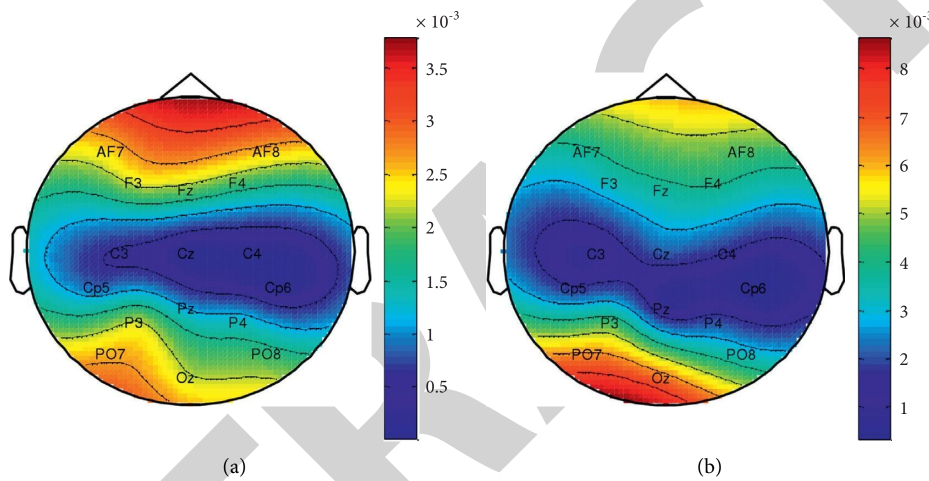


FIGURE 7: EEG correlation between positive and negative dichotomies of emotion. (a) Depressed emotion. (b) Anxiety emotion.

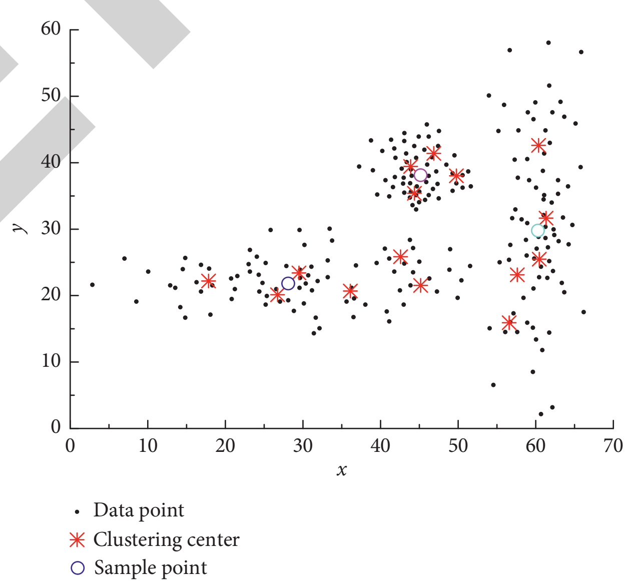


FIGURE 8: Two-dimensional view distribution of initial cluster center.

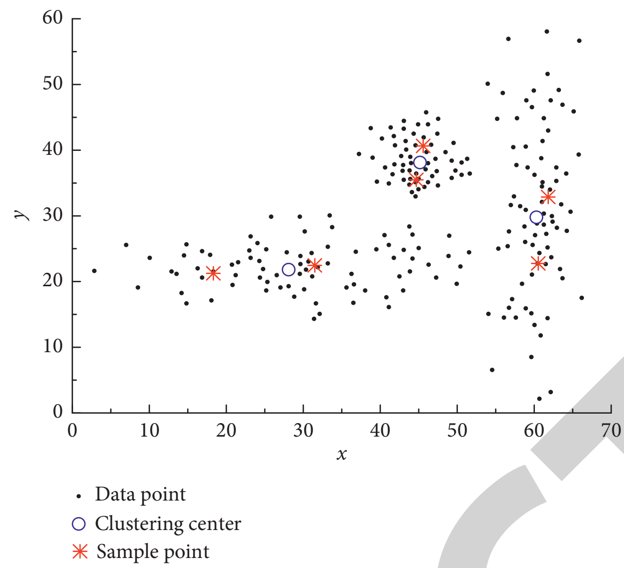


Figure 9: Clustering results after 9 iterations.

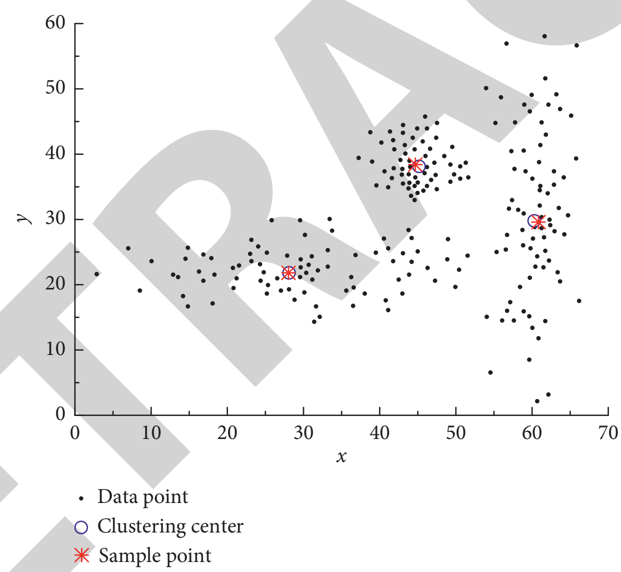


Figure 10: Clustering results after 17 iterations.

5.4. Comparative Analysis of Emotion Recognition. The proposed algorithm is compared with other algorithms to illustrate the effectiveness of the proposed algorithm. The classification of emotions here refers to the positive and negative (depression and anxiety) categories of tasks. The classification accuracy rate of the two emotional classifications is calculated and compared. This study was compared with DEAP BaseLine [28] and CNN [29]. The DEAP BaseLine is the baseline result of the DEAP database publisher's classification of emotions using traditional machine learning techniques. The average classification accuracy of different methods with the different number of features is shown in Figure 11.

As can be seen from Figure 11, regardless of the number of feature dimensions, the classification accuracy of the proposed method is significantly higher than the existing DEAP BaseLine and CNN methods, reaching the highest accuracy of 68.35%. The classification results of the four methods were ranked as ours, CNN, and DEAP BaseLine. This is because the combination of entropy-weighted clustering algorithm and sparse learning feature selection can extract the local correlation of heterogeneous data and obtain more discriminating abstract features, thus obtaining the higher accuracy of sentiment classification. When the feature dimension is about 300, the proposed method can achieve the highest average, while the classification results of other methods are almost unchanged. It should be noted that when the feature number is greater than 700, the classification accuracy of the proposed method is no longer changed. This is because all the

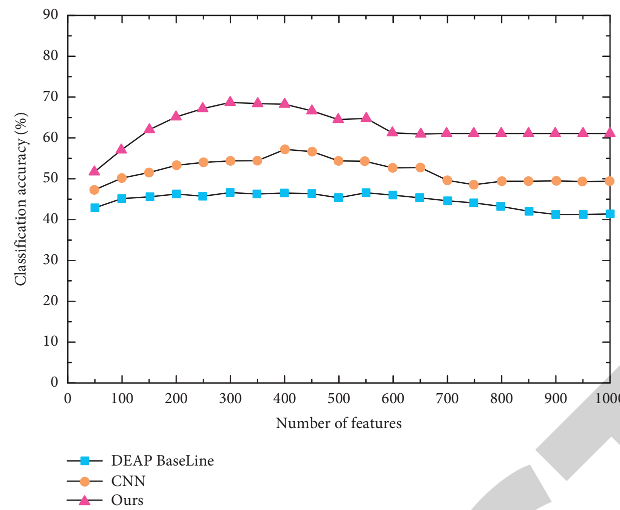


FIGURE 11: Average classification accuracy of different methods with different feature numbers.

features corresponding to the sparse weights have been used, and no more new features will be generated thereafter.

6. Conclusions

This paper proposes a method of college students' psychological emotion recognition based on sparse learning and entropy-weighted clustering. Sparse learning is mainly used to achieve the feature selection of EEG signals, which provides strong support for the subsequent construction of an emotional recognition model. At the same time, the entropy-weighted clustering algorithm and the features obtained after sparse learning are combined to establish a classification model for depression and anxiety, and better clustering accuracy and generalization are obtained. Under the condition of different feature dimensions, the classification accuracy of the proposed method is higher than that of the existing methods, and the highest accuracy reaches 68.35%. The experimental results of positive and negative dichotomies of emotions verify the effectiveness of the proposed method. In future research, we will focus on the finer spatiotemporal correlation extraction of electroencephalogram signals and try to identify emotions using nonlinear dynamic characteristics.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. J. Huster, S. Enriquez-Geppert, and C. F. Lavallee, "Electroencephalography of response inhibition tasks," *International Journal of Psychophysiology Official Journal of the International Organization of Psychophysiology*, vol. 87, no. 3, pp. 217–233, 2016.

[2] A. Reches, R. R. Nir, and M. J. Shram, "A novel electroencephalography-based tool for objective assessment of network dynamics activated by nociceptive stimuli," *European Journal of Pain*, vol. 20, no. 2, pp. 250–262, 2016.

[3] V. K. Kimiskidis, "Transcranial magnetic stimulation (TMS) coupled with electroencephalography (EEG): biomarker of the future," *Revue Neurologique*, vol. 20, no. 6, pp. 123–126, 2016.

[4] T. Sara, N. Chris, and P. Agnès, "Older people's experiences of mobility and mood in an urban environment: a mixed methods approach using electroencephalography (EEG) and interviews," *International Journal of Environmental Research and Public Health*, vol. 14, no. 2, pp. 151–157, 2017.

[5] T. Heunis, C. Aldrich, and V. P. J. De, "Recent advances in resting-state electroencephalography biomarkers for autism spectrum disorder—a review of methodological and clinical challenges," *Pediatric Neurology*, vol. 11, no. 3, pp. 28–37, 2016.

[6] B. Tobias and T. Myron, "Induction of spreading depression in the ischemic hemisphere following experimental middle cerebral artery occlusion: effect on infarct morphology," *Journal of Cerebral Blood Flow and Metabolism*, vol. 16, no. 2, pp. 202–213, 2016.

[7] U. J. Chaudhary, D. W. Carmichael, and R. Roman, "Mapping preictal and ictal haemodynamic networks using video-electroencephalography and functional imaging," *Brain*, vol. 135, no. 12, pp. 3645–3663, 2017.

[8] E. Albornoz and D. Milone, "Emotion recognition in never-seen languages using a novel ensemble method with emotion profiles," *IEEE Transactions on Affective Computing*, vol. 8, no. 99, pp. 21–29, 2017.

[9] J. C. Castillo, Á. Castro-González, and A. Fernández-Caballero, "Software architecture for smart emotion recognition and regulation of the ageing adult," *Cognitive Computation*, vol. 8, no. 2, pp. 1–11, 2016.

[10] A. Russo, P. Jayakar, and M. Lallas, "The diagnostic utility of 3D electroencephalography source imaging in pediatric epilepsy surgery," *Epilepsia*, vol. 57, no. 1, pp. 24–31, 2016.

[11] C. Jing, B. Hu, and W. Yue, "Subject-independent emotion recognition based on physiological signals: a three-stage

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9769165, 1 page
https://doi.org/10.1155/2023/9769165

Retraction

Retracted: An Improved Image Steganography Framework Based on Y Channel Information for Neural Style Transfer

Security and Communication Networks

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Lin, X. Zhu, W. Ye, C. Chang, Y. Liu, and C. Liu, "An Improved Image Steganography Framework Based on Y Channel Information for Neural Style Transfer," *Security and Communication Networks*, vol. 2022, Article ID 2641615, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2641615, 12 pages
https://doi.org/10.1155/2022/2641615

WILEY | Hindawi

Research Article

An Improved Image Steganography Framework Based on Y Channel Information for Neural Style Transfer

Wenjie Lin,[1] Xueke Zhu,[1] Wujian Ye,[1] Chin-Chen Chang,[2] Yijun Liu,[1] and Chengmin Liu[1]

[1]*School of Information Engineering, Guangdong University of Technology, Guangzhou, Guangdong 510006, China*
[2]*Department of Information Engineering and Computer Science, Feng Chia University, Taichung, Taiwan*

Correspondence should be addressed to Wujian Ye; yewjian@126.com and Chin-Chen Chang; alan3c@gmail.com

Received 7 December 2021; Revised 9 January 2022; Accepted 11 January 2022; Published 29 January 2022

Academic Editor: Chin-Ling Chen

Neural style transfer has effectively assisted artistic design in recent years, but it has also accelerated the tampering, synthesis, and dissemination of a large number of digital image resources without permission, resulting in a large number of copyright disputes. Image steganography can hide secret information in cover images to realize copyright protection, but the existing methods have poor robustness, which is hard to extract the original secret information from stylized steganographic (stego) images. To solve the above problem, we propose an improved image steganography framework for neural style transfer based on Y channel information and a novel structural loss, composed of an encoder, a style transfer network, and a decoder. By introducing a structural loss to restrain the process of network training, the encoder can embed the gray-scale secret image into Y channel of the cover image and then generate steganographic image, while the decoder can directly extract the above secret image from a stylized stego image output by the style transfer network. The experimental results demonstrate that the proposed method can effectively recover the original secret information from the stylized stego image, and the PSNR of the extracted secret image and the original secret image can reach 23.4 and 27.29 for the gray-scale secret image and binary image with the size of 256×256, respectively, maintaining most of the details and semantics. Therefore, the proposed method can not only preserve most of the secret information embedded in a stego image during the stylization process, but also help to further hide secret information and avoid steganographic attacks to a certain extent due to the stylization of a stego image, thus protecting secret information like copyright.

1. Introduction

Nowadays, deep learning technology has made remarkable breakthroughs in computer vision, text processing, clustering [1, 2], and voice recognition with its powerful ability of feature learning. As one of the hot deep learning applications, neural style transfer (NST) not only improves the efficiency and quality of stylization, but also enriches people's lives and brings convenience to art design. However, it also leads to more and more unauthorized digital images being arbitrarily tampered with and synthesized, which seriously damages the rights and interests of copyright owners and the development of related industries. Therefore, how to protect the original image from illegal use in the process of stylization is one of the urgent problems to be solved.

Image steganography is a technique for hiding confidential information from public information. Using the characteristics of image files, we can hide secret image that we want to deliberately hide or prove identity and copyright information into the cover images. The cover image containing secret information is referred to as steganographic (stego) image for normal use or transmission, and its visual effect is almost the same as the original cover image, which would hardly affect the application scene of the cover image [3]. Most traditional methods superimpose secret information directly on a specific area of the image, which may cause great damage to the original image. The emerging deep learning-based methods can make the secret information evenly distributed in the original cover image by automatically learning the deep features of the secret image and achieve more accurate steganography of the image. For example, the generative adaptive network (GAN) model is used to encode data into images with good results. However, these methods have poor robustness and are difficult to extract effective secret information from stylized stego images [4].

To protect and recover the secret information more effectively, we propose an improved image steganography framework adapting to neural style transfer, consisting of an encoder network, a style transfer network, and a decoder network. By introducing the Y channel information and a novel structural loss to restrain the training of decoder and encoder networks, the encoder can effectively embed the gray-scale secret image into the Y channel of the cover image, and the decoder can directly obtain the original secret information from the stylized stego image. The experimental results demonstrate that the proposed framework can realize better embedding and extraction of the gray-scale secret image before and after the style transfer. The PSNR between the decrypted secret image and the original secret image can reach 23.4, and most of the details and semantics are retained. Therefore, the proposed framework can not only basically maintain most of the hidden information in the stego image during the stylization process, but also better recover the secret image with almost the same visual effect from the stylized stego image. That is, steganography attacks can be better avoided and secret private information such as copyright can be protected to a certain extent.

The rest of this paper is organized as follows. Section 2 presents the state of the art of neural style transfer and deep learning-based image steganography technologies. Section 3 presents the proposed image steganography frameworks for neural style transfer in detail. Then, Section 4 carries out the experiment comparison and analysis. Section 5 gives the conclusion and future research direction.

2. Related Work

2.1. Neural Style Transfer. NST technology is one of the new applications of deep learning proposed by Gatys in 2016. It uses a convolutional neural network (CNN) to separate and fuse the features of a content image and a style image, so as to transfer the particular painting style to the natural content image and generate a novel artistic painting [5]. However, it took a long time due to its iterative processing. Johnson et al. [6] designed a fast style transfer framework based on a feed-forward neural network (FNN), which can train a corresponding stylization model for each style and obtain comparable quality artworks in real time.

To maintain image semantics, Gatys et al. [7] divided the style factors into space, color, brightness, and other perceptual information, and then combined the above information together to achieve effect control of style transfer. Sanakoyeu et al. [8] designed an encoder-decoder network with a style-aware loss function that can generate high-definition stylized images with diverse textures. Liu et al. [9] proposed a depth-aware loss to constrain the training of stylization model, so that the spatial depth features of generated image were consistent with the content image. Luan et al. [10] presented a novel high-fidelity style transfer method. By introducing regularization terms to constrain the change range of style transfer, the satisfying real image style transformation can be achieved in various scenes such as delayed transformation, weather, season, art style, and so on. Liu et al. [11] retained salient semantic and emotional information of the original image in combination with saliency prediction technology. To deal with multistyle transfer, Dumoulin et al. [12] realized multistyle transfer on a single network model, which can simultaneously learn and integrate 32 different styles, thus relieving the problem of large space consumption. Chen et al. [13] proposed an arbitrary style transfer method named StyleBank, to separate content and style by building a forward encoder-decoder network. Tian et al. [14] implemented a patch-level network model to achieve multistyle transfer by reconstructing image in style-swap layer. Huang et al. [15] realized the fast arbitrary stylization by introducing a novel adaptive instance normalization (AdaIN) layer. Li et al. [16] designed a WCT (white-color transform) layer to train different decoders to achieve arbitrary and real-time diversified stylization.

2.2. Deep Learning-Based Image Steganography. Compared with traditional methods, image steganography based on deep learning has attracted extensive attention due to its improved performance and more flexible application scenarios [17, 18]. Qian firstly proposed a novel Gaussian neuron CNN (GNCNN) model in 2015 [19] for secret data hiding, which processed the image by using KV kernel of traditional algorithm and CNN network. Xu et al. [20] added a deeper network model based on GNCNN and different activation functions to significantly improve the hiding effect. To better avoid steganography, Jian et al. [21] presented a new CNN model that incorporates high-pass filters and truncated linear unit (TLU) activation functions into the spatially rich model (SRM). Their model showed good performance because it can pass most steganography tests.

Baluja [22] proposed a deep encoder-decoder model using CNN network, which can embed a secret color image into a cover image of the same size. Volkhonskiy et al. [23] constructed an improved deep convolutional generative adversarial network (DCGAN) model to avoid steganography detection and improve the capacity of embedding. Zhu et al. [24] also trained an encoder-decoder network combined with a novel noise layer (called as HiDDeN), which can well hide and extract secret information and generate a stego image that is visually indistinguishable from the cover image. Inspired by QR code, Tancik et al. [25] proposed a hyperlink embedding and extraction technique based on an encoder-decoder model. The encoder realized the hidden embedding of hyperlinks, and the decoder can be loaded into hardware like mobile phones to extract the hidden hyperlinks by taking photos of stego image. Image steganography is also used in combination with style transfer techniques in various scenarios [26]. Chen et al. [27] proposed a two-stage model and an end-to-end model for retaining the input content image information during style transfer process through steganography technology, effectively eliminating artifacts introduced in the reconstruction of stylized images. Zhong et al. [28] designed an image steganography technology combined with style transfer, which generated two different stylized images for content images, one for steganographic embedding and the other for steganographic analysis comparison, so as to improve the resistance of model steganographic analysis. To further improve the antisteganography algorithm, Li et al. [29] proposed an image steganography method based on style transfer and quaternion exponent moments (QEM). The arithmetic invariance of QEM is used to embed the secret information, and then, the stego image is style transferred. Finally, the secret image is extracted while the content image is reconstructed. Zhang et al. [30] proposed an image conversion network CSST-Net based on arbitrary style transfer, which can constrain style transfer by encoding secret information into adaptive steganography matrix, while directly synthesizing stylized images embedded with secret information during style transfer. Although the embedded information of deep learning-based method is still not high enough, its appearance greatly enriched the application scenarios of image steganography, which reflected that deep learning could not be limited to improving the embedding capacity and quality, and could have strong robustness.

2.3. Analysis and Summary. Most existing stylization methods tend to improve the efficiency of stylization and the quality of artworks, but their lack of consideration to effectively protect images involved in style transfer increases the risks of these images being arbitrarily modified, integrated, and disseminated, which make it difficult to protect the legal rights of the copyright owner. The deep learning-based image steganography technique extends the types of secret information that can be embedded, not only binary-coded character information, but also any image information. Most of the existing methods can embed secret images well, but they cannot effectively extract secret information from stylized stego images, because the structure of the original stego image is seriously disturbed or even lost in the process of style transfer, which ultimately leads to the inability to maintain the hidden secret information.

3. The Improved Image Steganography Framework for Neural Style Transfer

As the above analysis, the existing stylization methods fail to provide effective protection for the secret information of original content images. At the same time, the existing steganographic methods have poor robustness to stylized stego images. To solve the above problem, we propose an improved image steganography framework adapting to neural style transfer by introducing Y channel information and a structural loss. Figure 1 shows the structure of the proposed framework including an encoder network, a style transfer network, and a decoder network.

The encoder network generates a stego image by embedding a secret image into Y channel of a cover image; the decoder network extracts the secret image from the stylized stego image. Similar to the model in [18], the architecture of encoder/decoder network is composed of five convolutional layers with 50 convolution kernels of $\{3 \times 3, 4 \times 4, 5 \times 5\}$, as shown in Figure 2.

The neural style transfer network shown in Figure 3 transfers specific style to the stego image, which is a fast image transformation network used in [11], consisting of 3 convolutional layers, 4 residual blocks, 2 deconvolutional layers, and a final output layer. And the tanh function is adopted to constrain the pixel values of an output image to be between 0 and 255.

We use a novel structural loss and a perceptual loss to train the above three networks as a whole to hide and recover secret information efficiently under the situation of style transfer and improve the robustness of image steganography model. Specifically, the above two kinds of losses are applicable to constrain the training of the encoder and the decoder, and the transfer network is a pretraining model used to generate stylized stego images as the input of decoder.

3.1. Perception Loss. According to [11], the Y channel contains rich semantic information. Thus, to improve the image quality after embedding secret image I_s, we propose to embed the I_s into the Y channel of a cover image I_c to obtain a stego image I_{s+c}; then, the stylized stego image I_{s+c}^s is obtained by the style transfer network with the input of I_{s+c}; at last, the hidden secret image I_s^o is decoded by the decoder network with the input of I_{s+c}^s.

So as to constrain the network training process and keep the stego image similar to cover image encoded by encoder network, and the original secret image similar to the decrypted image decoded by the decoder network, we take the weighted sum of 1-norms of the original secret image and the decrypted image, and the weighted sum of 1-norms of the cover image and the secret image as the final perceptual loss. The calculation of perceptual loss is shown as follows:

$$\ell_{\mathrm{perc}}\left(I_c, I_{s+c}, I_s, I_s^o\right) = \alpha I_c - I_{s+c} + \beta I_s - I_s^o, \qquad (1)$$

where I_c, I_{s+c}, I_s and I_s^o refer to the cover image, the stego image, the original secret image, and the decrypted secret image obtained by the decoder, respectively, α and β are the weight parameters, and ℓ_{perc} refers to the perceptual loss.

3.2. Structural Loss. The larger the SSIM value (between 0 and 1), the more similar the structure of the two images. Similar to the literature [31], SSIM value is used to measure the structure similarity of two images. Thus, a structural loss is constructed by calculating the SSIM value between the original secret image I_s and the decrypted secret image I_s^o, which can better keep the structure information of the secret image. The calculation of structural loss is shown as follows:

$$\ell_{\mathrm{stru}} = \ell_{\mathrm{stru}}\left(I_s, I_s^o\right) = \sigma\phi\left(I_s\right) - \phi\left(I_s^o\right)_2^2 + \rho\left(1 - \mathrm{SSIM}\left(I_s, I_s^o\right)\right), \qquad (2)$$

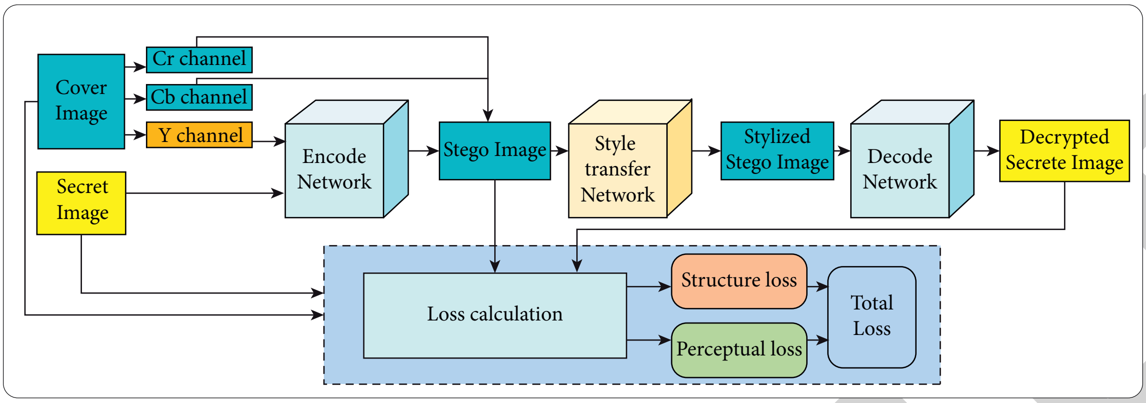


Figure 1: The structure of the proposed framework.

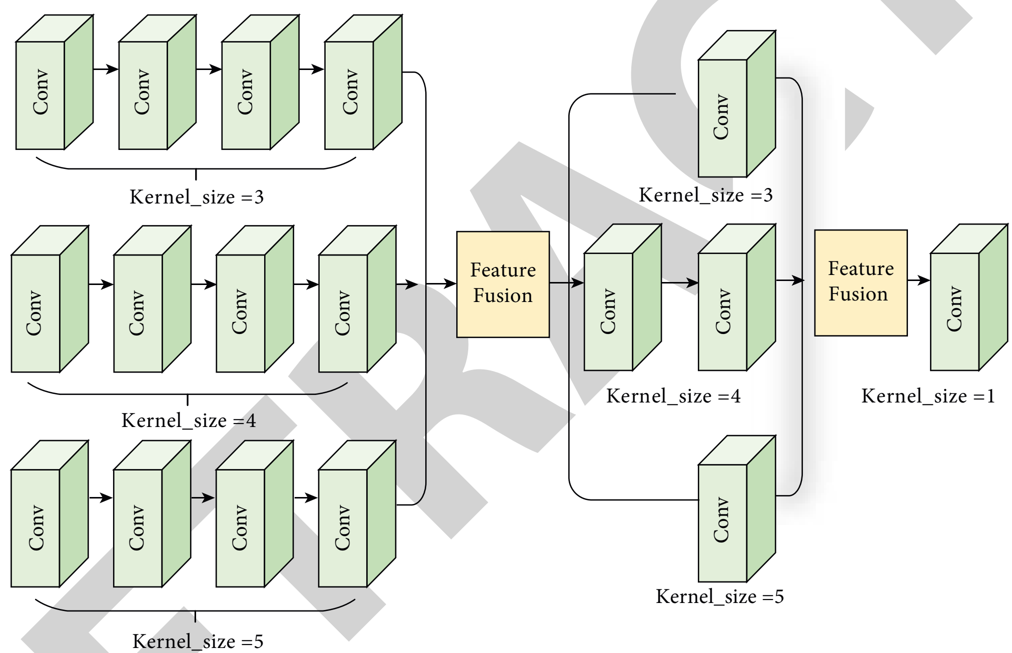


Figure 2: The structure of encoder/decoder network.

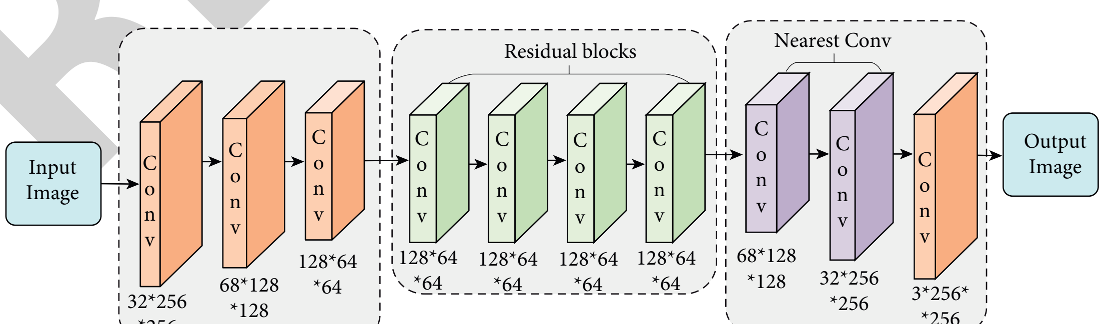


Figure 3: Neural style transfer network.

where $\phi(I_s)$ and $\phi(I_s^o)$ are the pixel matrix of the original secret image and the decrypted secret image, and σ and ρ are the weight parameters, respectively.

3.3. Total Loss. The total loss of the proposed network consists of cover loss ℓ_{cover} and secret loss ℓ_{secret}, and both of them are equal to the weighted sum of perception loss and structural loss, calculated as follows, where γ and δ are adjustable weight values.

$$\ell_{\text{cover}} \text{ or } \ell_{\text{secret}} = \gamma \ell_{\text{perc}} + \delta \ell_{\text{stru}}. \tag{3}$$

Finally, the weighted sum of the cover loss ℓ_{cover} and secret loss ℓ_{secret} is calculated as the total loss ℓ_{total} of the overall network shown in the following equation, where λ and η are the weight parameters of the cover loss and the secret loss.

$$\ell_{\text{total}} = \lambda \ell_{\text{cover}} + \eta \ell_{\text{secret}}. \tag{4}$$

4. The Experimental Analysis

In this paper, we design the training process of the proposed method shown in Figure 1. In each forward calculation, a cover image and a secret image are preprocessed and input to the encoder network. After the encoder network completes the embedding of secret image, stego image is output. The stego image is input into the pretrained style transfer network to obtain the stylized stego image, which can be processed by the decoder network to decode the embedded secret image. At this point, the corresponding perceptual loss and structural loss are calculated. In back propagation, the proposed method carries out back propagation according to the loss value and updates the parameters of the encoder and the decoder network. After many times of forward calculation and back propagation, the optimized encoder and decoder are obtained.

4.1. Platform and Dataset. In the experiments, 53228 images are randomly selected from Microsoft CoCo2017 [32], among which 26614 images are used as cover images, and the rest is secret images for model training. All images are resized into 256×256 pixels before training.

In the experiment, the graphics card is NVIDIA Tesla P40 GPU with a memory of 16 GB, and the CPU is Intel(R) Xeon(R) CPU E5-2620 v4 with memory of 128G. And the deep learning software is PyTorch 0.4.0 with CentOS 7 operating system and Python 3.7 platform. The learning rate is set to 1×10^{-3}, and the Adam function is used to optimize the network. For the weight parameters, we set the $\alpha\colon\beta$ in Equation (1) to 1 : 1, the $\sigma\colon\rho$ in Equation (2) to 3 : 10, the $\gamma\colon\delta$ of cover loss and secret loss in Equation (3) to 7 : 10 and 1 : 1, and the $\lambda\colon\eta$ in Equation (4) to 2 : 1, according to our experience.

Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as the evaluation indexes. PSNR is adopted to evaluate the quality of a compressed image compared with the original image, and the higher the value, the less distortion in the compressed image. SSIM is usually used for measuring image similarity in terms of three aspects of structure, brightness, and contrast [11]. The larger the value, the more similar the structure of the two images.

4.2. The Color Image and the Gray-Scale Image Regarded as Secret Images, Respectively. To explore the robustness of encoder-decoder model proposed by Baluja et al. [20] for neural style transfer, we use color images as secret images in this subsection. As shown in Figure 4, the third and fifth rows are original stego images and stylized stego images, and the fourth and sixth rows are the corresponding extracted secret images.

As observed in Figure 4, the original secret images can be completely extracted from the stego images by the decoder model, but they can hardly be decoded from the stylized stego images, which are almost distorted with basically cluttered noise point composition. This is because the color image has a large amount of information, and the stylization process destroys more embedded information, which makes it difficult to decode effective information from the stylized image. Thus, Baluja's method is not effective enough and has poor robustness in the scenario of neural style transfer.

4.3. Gray-Scale Image and Y Channel Information Embedding. After style transfer, the structure and color of steganographic images are fused and updated accordingly with style features of style image. Therefore, the original semantic information of stego images is damaged to varying degrees, leading to the disruption and destruction of the embedded secret information.

The rapid style transfer network has great ability of semantic preservation, which shows that the Y channel of images contains more semantic information [23]. Thus, we compare the extraction ability of two different embedding schemes for embedding secret images. Scheme 1 embeds a gray-scale image into the cover image. Scheme 2 encodes a gray-scale image into the Y channel of the cover image.

As shown in Figure 5, both schemes can embed secret gray-scale images well and generate stego images with similar quality, and their stylized effects are also good. As observed from Scheme 1, the decoder can basically extract the outline information of the secret images from the stylized stego images, but most of the detailed contents and textures are still lost, and the extraction quality still needs to be improved. In Scheme 2, the decoder can more clearly decode the main content information of the secret image from the stylized stego image. For example, the edges and textures such as faces, clothes, umbrellas, and road signs in the images can be clearly distinguished. Compared with Scheme 1, Scheme 2 has better steganography performance and greatly reduced the distortion of the secret information.

4.4. Combining Y Channel Embedding and Structural Loss to Optimize the Decoder Model. To further improve the clarity of the decrypted secret image, we construct a structural loss to constrain the decoder model training. The first, second, third, and fourth columns are the gray-scale secret image, cover image, stego image, and stylized stego image, respectively; the fifth, sixth, and seventh columns are three different schemes proposed in this paper, in which Scheme 1 and Scheme 2 are in subsection 4.3.2 and Scheme 3 proposed in this subsection.

As shown in Figure 6, the quality of the secret image extracted by Scheme 3 is significantly improved compared to Scheme 2, and most of the detailed texture and structures in

Figure 4: The effects of Baluja's method using color images.

the images are maintained, which is not visually distinguishable from the original image. For example, the edges of umbrellas and clothes, and the markings of giraffes are smoother and clearer, without obvious mosaics. In addition, it can be seen from the residual images of the secret images extracted by each method and the original secret images that the extraction quality of Scheme 1 is the worst, and most of the information is wrong. Scheme 3 can recover most information better with less error bits.

To quantitatively analyze the experimental results, we randomly select 30 pairs of secret images and cover images and use two different style models to evaluate the proposed methods in this paper. The PSNR and SSIM values between the cover image and the corresponding stego image, and the secret images and the decrypted images are calculated, respectively. Then, the mean values of PSNR and SSIM of each scheme are calculated as shown in Table 1.

As observed in Table 1, Scheme 3 better balances the quality between the stego image and the decrypted secret image, which also can make the distortion of the stego image relatively small. The other two schemes can produce stego images with better quality because the secret images are directly embedded in the cover images without constraint. However, due to the stylization processing, the visual gap of stego images generated by various methods is not too obvious, which will not affect its subsequent use. Furthermore, the effect of secret images extracted by them is poor, which is difficult to be applied in actual scenes.

Figure 7 shows the total loss curve, secret loss curve, and cover loss curve of the network model in this paper in training process. The size of secret image and cover image is 256×256. The total loss curve reflects that the network model is close to convergence during the 21st epoch iteration training. As can be seen from the figure, encoder network loss and decoder network loss, which constitute the total loss, also converge at this point, and their convergence values are different due to the different weighted weights of them.

4.5. Comparison and Analysis of Existing Methods. In this subsection, MNIST dataset is used as secret images, and the influence of the size of embedded secret images on model's

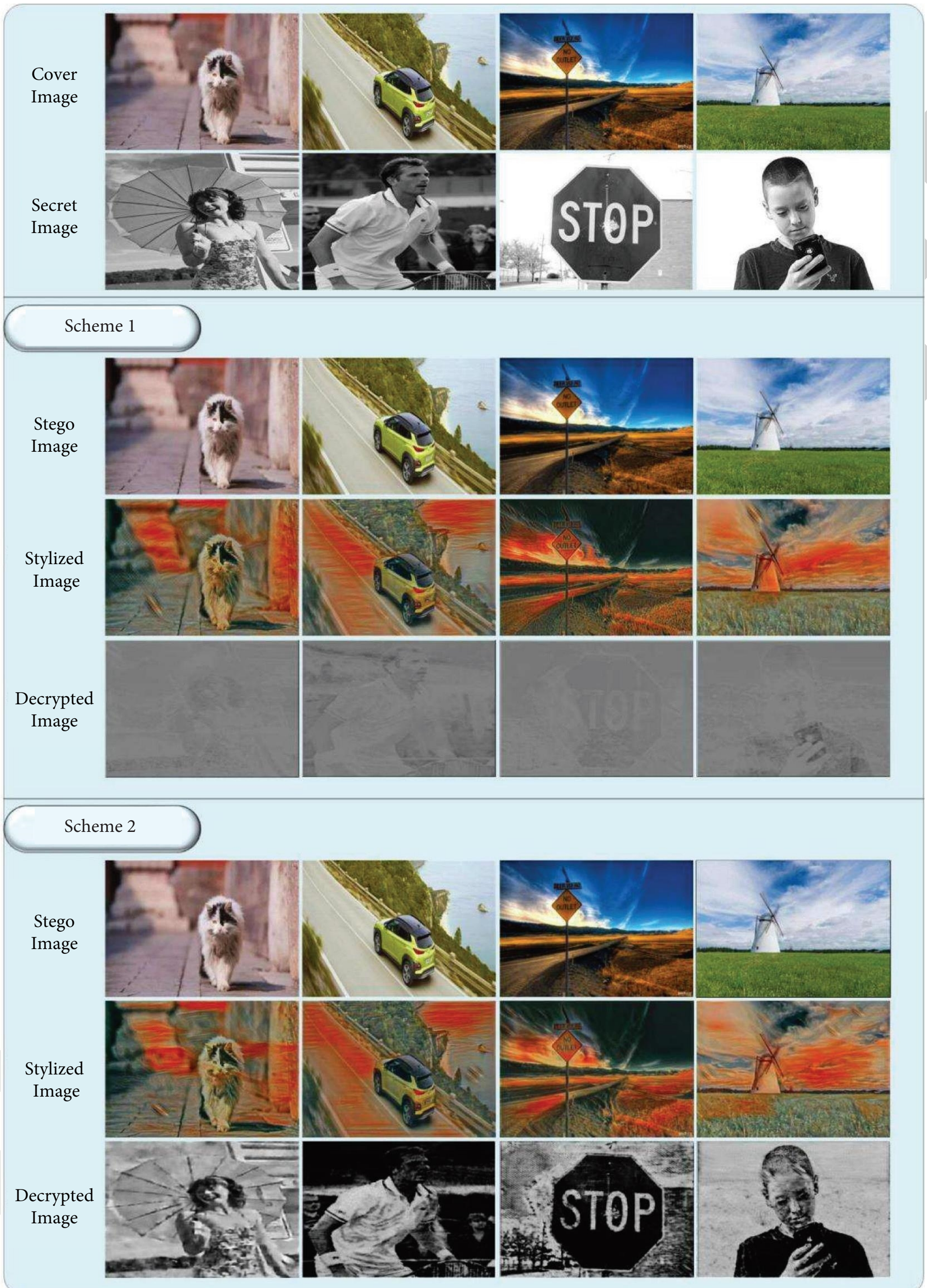


Figure 5: The extracted effects by different schemes.

results is analyzed and compared with Li's method [26]. Specifically, we embedded and extracted 50 secret images of different sizes, including 256×256, 128×128, 32×32, and 28×28, respectively. The size of the cover image is 256×256. Figure 8 shows the corresponding effects.

As observed from Figure 8, as the size of the embedded secret image decreases, the encoder can better embed the secret image into the cover image, the visual effect of the stylized stego image is also good, and the secret image extracted by the decoder is getting better and better.

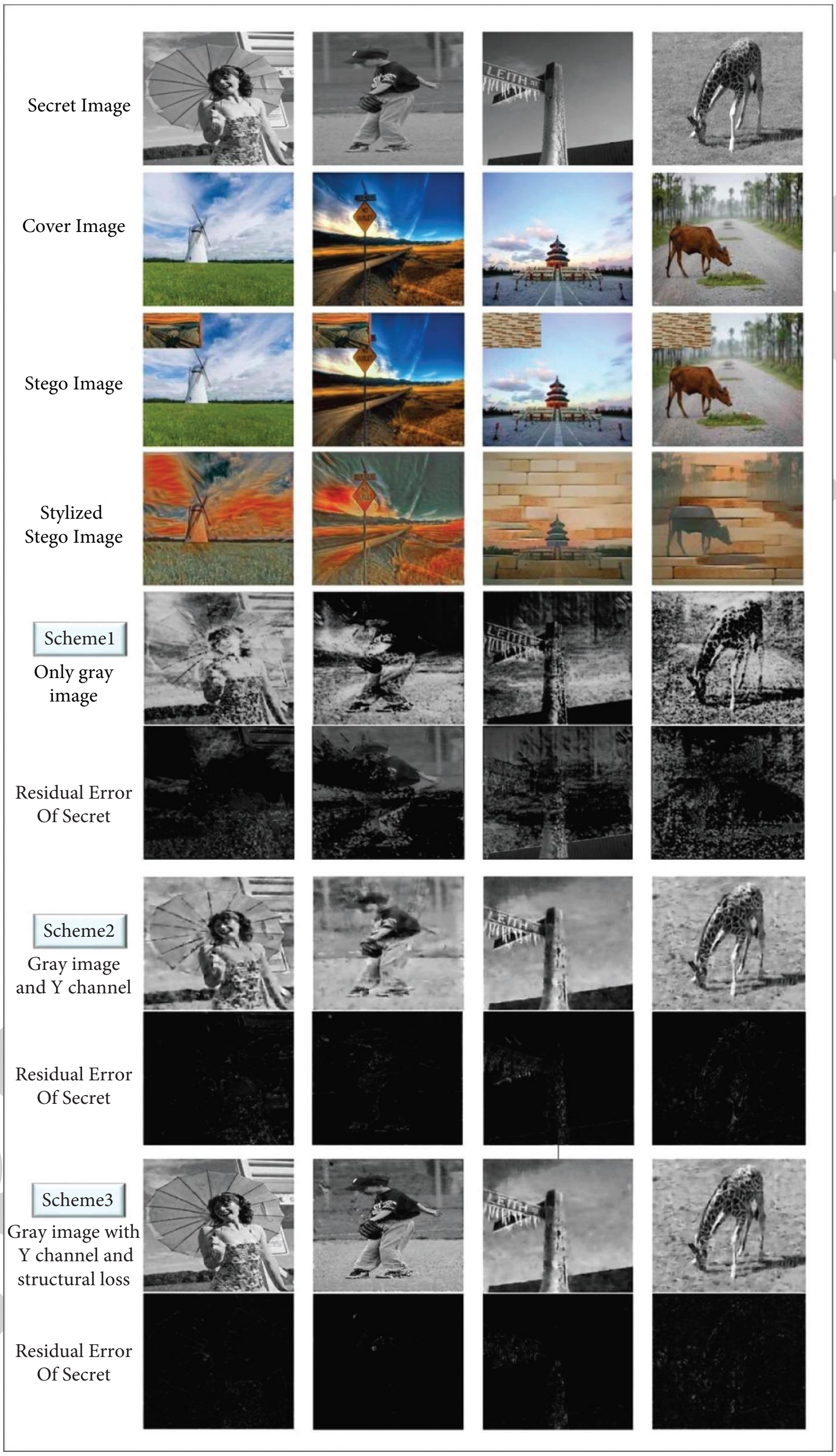


FIGURE 6: The visual effects of secret images extracted by different proposed schemes.

Table 1: The mean values of PSNR and SSIM of each scheme.

Evaluation indexes		Scheme 1 (only gray-scale image)	Scheme 2 (gray-scale image and Y channel)	Scheme 3 (gray-sale image with Y channel and structural loss)
Between original secret images and extracted images	PSNR	11.4829	13.0054	**24.1432**
	SSIM	0.5438	0.7529	**0.8027**
Between cover images and stego images	PSNR	**31.7026**	31.3511	26.0245
	SSIM	**0.8976**	0.8150	0.7850

The bold fonts indicate that the corresponding schemes have the highest PSNR and SSIM values, that is, the secret images extracted by Scheme 3 are most similar to the original images. And the values between cover images and stego images of Scheme 1 is the highest, but the visual effects of three schemes are acceptable since the stylized stego images are almost the same to each other.

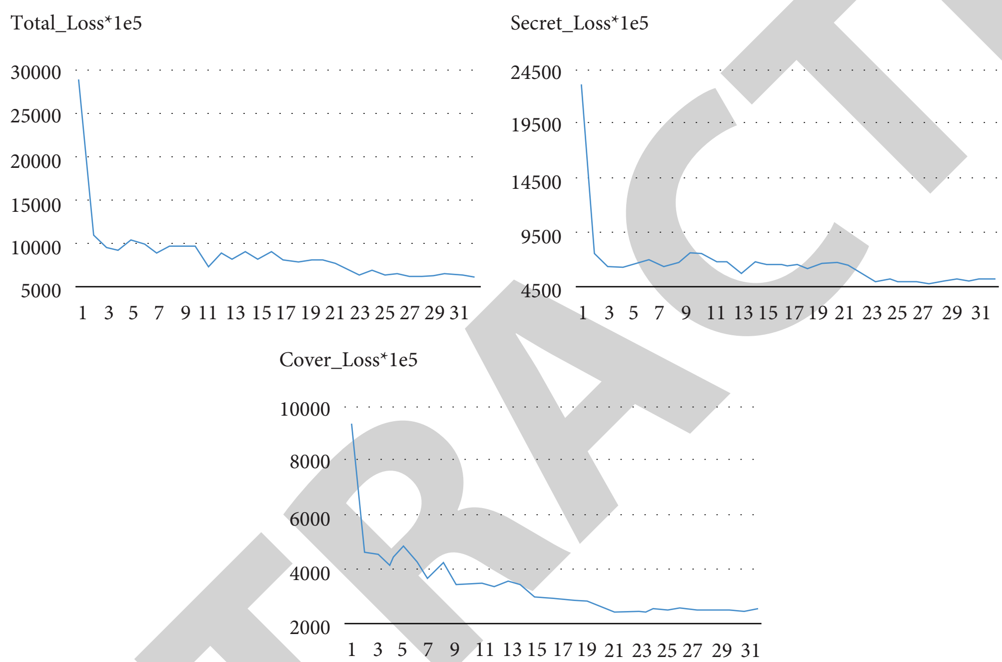


Figure 7: Convergence of loss curve. (a) Total loss. (b) Secret loss. (c) Cover loss.

However, with the increase of the size of the embedded secret image, more noises appear in the decrypted secret image.

To make quantitative analysis of the above effects, we calculate the PSNR and SSIM between original secret image and decrypted secret image with different sizes, and the results are shown in Table 2. The effect of the method proposed in this paper is far better than that of Li's method [29], the PSNR is 31.35, SSIM is 0.96, and BER is 0.06 in case of secret images of 32×32. This is because the proposed method can directly extract the secret image from the stylized stego image, while Li's method needs to extract the secret image from the destylized stego image relying on training an additional destylization network, thus losing more original secret information.

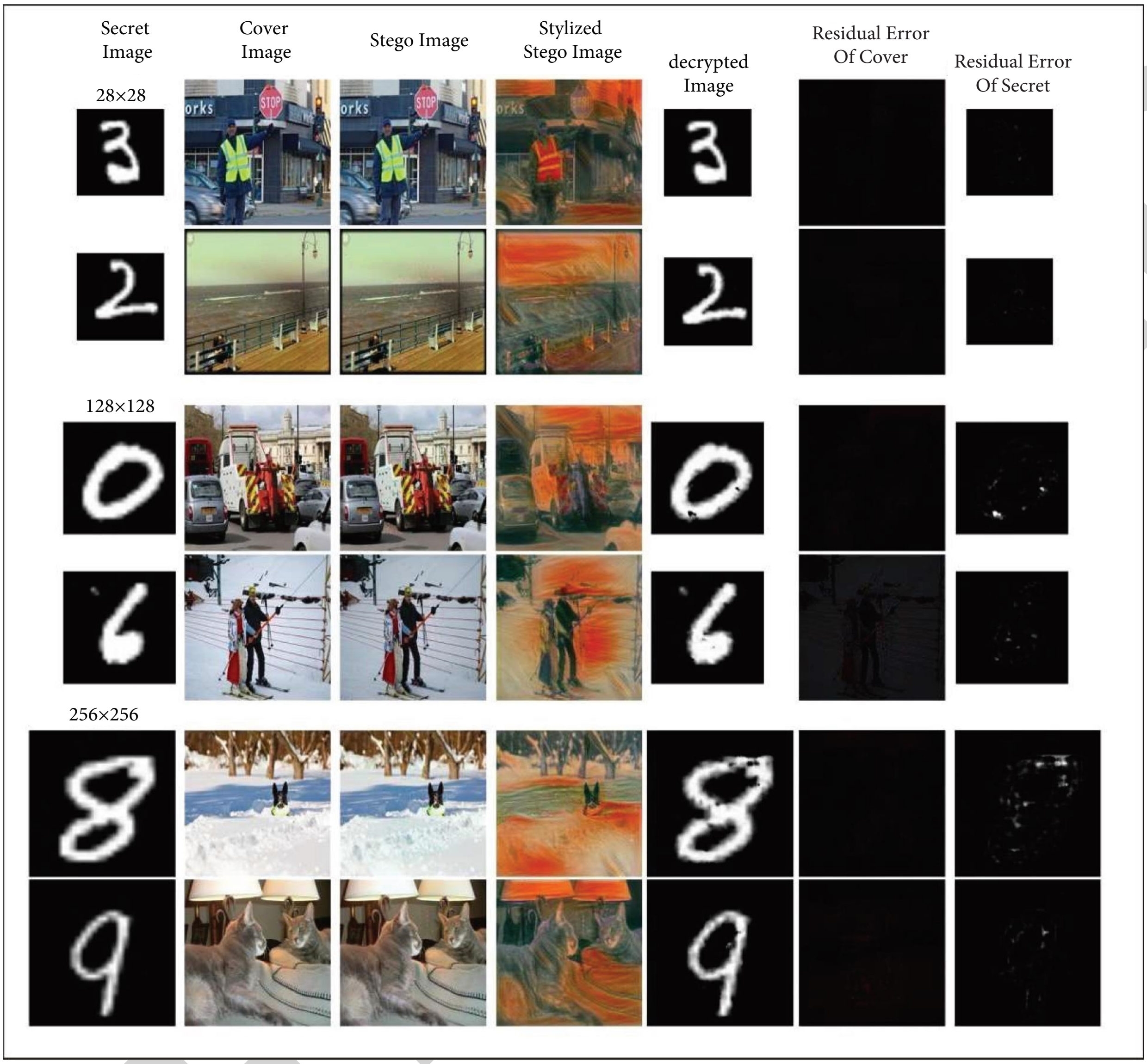


Figure 8: The visual effects for embedding secret images of different sizes.

Table 2: The mean PSNR and SSIM values of cover images and secret images.

Size of cover image	Size of secret image	Secret image PSNR		SSIM		BER	
		Our	Li [29]	Our	Li [29]	Our	Li [29]
256×256	28×28	31.35	—	0.96	—	0.04	—
	32×32	30.21	—	0.94	—	0.06	0.09
	128×128	27.43	—	0.85	—	0.17	—
	256×256	27.29	—	0.82	—	0.22	—

5. Conclusion and Future Works

In this paper, we attempt to explore the combination of image steganography and style transfer techniques, and to study how to effectively maintain the secret information of the stego image (referred to the content image) before and after stylization, so as to achieve better protection of private information like copyright. Thus, we proposed a novel steganography framework for neural style transfer by introducing the Y channel information and a structural loss, which enables the secret information to be efficiently embedded into the cover image and directly recovered from stego images or stylized stego images with high quality. The experimental results show that the proposed method has high robustness to the stylization and expands the application scenarios of image steganography.

However, this technology still has a large space for optimization. For instance, the structure of encoder-decoder network can be further optimized to improve the image quality of encrypted and decrypted images. The capacity of embedded secret information needs to be increased, so that the model can embed the color images and then can decode the original secret information with high quality from stego images processed or disturbed by different techniques. That is, we can expand the application of image steganography technology in different situations, such as image conversion, coloring, generation, etc., to achieve more expansion and improvement.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by Key-Area Research and Development Program of Guangdong Province under Grant nos. 2018B030338001, 2018B010115002, and 2018B010107003 and in part by Innovative Talents Program of Guangdong Education Department and Young Hundred Talents Project of Guangdong University of Technology under Grant no. 220413548.

References

[1] X. Yang, G. Lin, Y. Liu, F. Nie, and L. Lin, "Fast spectral embedded clustering based on structured graph learning for large-scale hyperspectral image," *IEEE Geoscience and Remote Sensing Letters*, vol. 99, pp. 1–5, 2020.

[2] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review*, vol. 40, no. 1, Article ID 100379, 2021.

[3] J. Fridrich, M. Goljan, and D. Rui Du, "Detecting LSB steganography in color, and gray-scale images," *IEEE Multimedia*, vol. 8, no. 4, pp. 22–28, 2001.

[4] A. A. Tamimi, A. M. Abdalla, and O. Alallaf, "Hiding an image inside another image using variable-rate steganography," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 10, pp. 1–4, 2013.

[5] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423, Las Vegas, NV, USA, June 2016.

[6] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proceedings of the European Conference on Computer Vision*, pp. 694–711, Computer Vision - ECCV 2016, Amsterdam, The Netherlands, October 2016.

[7] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, "Controlling perceptual factors in neural style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3730–3738, Honolulu, HI, USA, July 2017.

[8] A. Sanakoyeu, D. Kotovenko, S. Lang, and B. Ommer, "A style-aware content loss for real-time HD style transfer," in *Proceedings of the European Conference on Computer Vision*, pp. 715–731, Computer Vision - ECCV 2018, Germany, September 8-14, 2018.

[9] X. Liu, M. Cheng, Y. Lai, and P. Rosin, "Depth-aware neural style transfer," in *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*, pp. 1–10, Los Angeles California, July 2017.

[10] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Honolulu, HI, USA, July 2017.

[11] Y. Liu, Z. Xu, W. Ye et al., "Image neural style transfer with preserving the salient regions," *IEEE Access*, vol. 7, Article ID 40037, 2019.

[12] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," in *Proceedings of the Conference on ICLR*, pp. 1–26, Toulon, France, April 2017.

[13] D. Chen, L. Yuan, J. Liao, N. Yu, and H. G. StyleBank, "An explicit representation for neural image style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–10, CVPR), Honolulu, HI, USA, July 2017.

[14] Q. Tian and M. Schmidt, "Fast patch-based style transfer of arbitrary style," in *Proceedings of the Conference on Neural Information Processing Systems*, pp. 1–5, Barcelona, Spain, December 2016.

[15] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1510–1519, Venice, Italy, October 2017.

[16] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M. Yang, "Universal style transfer via feature transforms," in *Proceedings of the Conference on Neural Information Processing Systems*, pp. 1–11, Long Beach, United States, December 2017.

[17] J. Qin, Y. Luo, X. Xiang, Y. Tan, and H. Huang, "Coverless image steganography: a survey," *IEEE Access*, vol. 7, no. 99, 171394 pages, Article ID 171372, 2019.

[18] M. Dalal and M. Juneja, "A secure video steganography scheme using DWT based on object tracking," *Information Security Journal: A Global Perspective*, no. 1, pp. 1–18, 2021.

[19] Y. Qian, D. Jing, W. Wei, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Proceedings of the SPIE-International Society for Optical Engineering*, vol. 9409, pp. 1–10, Article ID 94110J, San Francisco, CA, United States, March 2015.

[20] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 708–712, 2016.

[21] Y. Jian, J. Ni, and Y. Yang, "Deep learning hierarchical representations for image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2545–2557, 2017.

[22] S. Baluja, "Hiding images in plain sight: deep steganography," *Advances in Neural Information Processing Systems*, vol. 30, pp. 2069–2079, 2017.

[23] D. Volkhonskiy, I. Nazarov, and E. Burnaev, "Steganographic generative adversarial networks," in *Proceedings of the Conference on Neural Information Processing Systems*, pp. 1–8, Long Beach, United States, December 2017.

[24] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei, "HiDDeN: hiding data with deep networks," in *Proceedings of the European

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9895458, 1 page
https://doi.org/10.1155/2023/9895458

Retraction

Retracted: Multipurpose Watermarking Approach for Copyright and Integrity of Steganographic Autoencoder Models

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. Gu, C. Chang, Y. Bai, Y. Fan, L. Tao, and L. Li, "Multipurpose Watermarking Approach for Copyright and Integrity of Steganographic Autoencoder Models," *Security and Communication Networks*, vol. 2021, Article ID 9936661, 12 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2021, Article ID 9936661, 12 pages
https://doi.org/10.1155/2021/9936661

Research Article

Multipurpose Watermarking Approach for Copyright and Integrity of Steganographic Autoencoder Models

Wei Gu,[1] **Ching-Chun Chang**,[2] **Yu Bai**,[3] **Yunyuan Fan**,[3] **Liang Tao**,[1] **and Li Li**[3]

[1]*School of Computer Science and Technology, Anhui University, Hefei 230039, China*
[2]*Department of Computer Science, University of Warwick, Coventry CV47AL, UK*
[3]*School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China*

Correspondence should be addressed to Li Li; lili2008@hdu.edu.cn

Received 13 November 2021; Accepted 13 December 2021; Published 30 December 2021

Academic Editor: Chin-Ling Chen

With the great achievements of deep learning technology, neural network models have emerged as a new type of intellectual property. Neural network models' design and training require considerable computational resources and time. Watermarking is a potential solution for achieving copyright protection and integrity of neural network models without excessively compromising the models' accuracy and stability. In this work, we develop a multipurpose watermarking method for securing the copyright and integrity of a steganographic autoencoder referred to as "HiDDen." This autoencoder model is used to hide different kinds of watermark messages in digital images. Copyright information is embedded with imperceptibly modified model parameters, and integrity is verified by embedding the Hash value generated from the model parameters. Experimental results show that the proposed multipurpose watermarking method can reliably identify copyright ownership and localize tampered parts of the model parameters. Furthermore, the accuracy and robustness of the autoencoder model are perfectly preserved.

1. Introduction

The latest achievements in deep learning (DL) have gained remarkable success in a number of fields [1], such as speech recognition [2, 3], visual computing [4, 5], and natural language processing [6, 7]. DL methods have been reported to outperform traditional methods substantially [6–10].

The production of a deep neural network model is remarkably costly, requiring a great quantity of training data and consuming massive amounts of computing resources and time. If the deep neural network model is maliciously copied, transmitted, or stolen, then the owner will suffer a terrible loss. Therefore, it is crucial to prevent the copyright and integrity of such intellectual property (IP) from being violated. The recent development of various watermarking methods has triggered research attention in addressing the IP issues over DL models [11–13].

The following real-world application scenario is considered. For example, an organization has developed a product based on DL technology and put it into the market to achieve profitability. This action of the organization indicates that the purchaser of the product has the right to use the service within the scope allowed by law. However, if the customer uses this product for commercial purposes or provides it to other organizations, such use will be considered a serious violation. So, protecting the IP of the product is a difficult problem that must be solved in this scenario.

Some previous works [14–18] applied DL in many watermarking systems for images, videos, and audios to achieve better experimental results. However, rather than the used DL model, these works aim to protect multimedia copyright information. This condition motivated the current investigation regarding the IP protection of DL models.

First, Uchida et al. [14] and Nagai et al. [19] proposed a generic watermark embedding framework based on deep neural networks (DNNs) using a parametric regularizer; thus they could embed watermarks in the training phase of the model. Wang et al. [20] extended the work of Uchida et al. by adding a separate neural network to form a relationship mapping between the network weights and the watermark information. However, such an improvement cannot withstand the ambiguity attacks. To solve this problem, Rouhani et al. [21] proposed an end-to-end IP protection framework: DeepSigns that allows developers to insert watermarking information systems into DL models before distributing models. Fan et al. [22] applied their proposed DNN copyright verification algorithm for anti-forgery authentication about passports. This technique remains robust after the network is modified, especially for DNN ambiguity attacks. These articles mainly discussed the issue of IP certification through watermarking DNNs in the extensively used white-box scenario. The accuracy of the watermark model remains unaffected. However, it is necessary to know all the DNN parameters to extract the watermark information during ownership verification of the DL model. The white-box technique restricts its universal use in any scenario.

IP protection in black-box scenarios is proposed in [15, 16, 23–27]. Compared with the white-box technique, the black-box watermarking methods are suitable for DNN model protection. The DNN model should be able to provide API services during its ownership verification; this model can also withstand statistical attacks [15, 16].

Adi et al. [4] selected hundreds of abstract images and attached labels as a trigger set and simultaneously utilized it with other training sets to train the classification neural network. Zhang et al. [25] proposed that watermarks embedding can be achieved in conjunction with a remote verification mechanism. Next, they designed an algorithm that can identify the ownership of DL models, which in turn can be trained while learning user-exclusive watermarks. Finally, they executed prespecified predictions when observing watermark modes at inference. Zhao et al. [26] proposed a watermarking framework for GNNs, in which an indeterminate figure related to features and labels is initialized as the trigger input. By training the main GNN model with the trigger figure, the watermark can be distinguished from its result during certification. Wu et al. [28] introduced a novel digital watermarking framework suitable for deep neural networks that output images as a result. All the output of the images from a watermarking DNN in this framework will contain an exclusive watermark. The basic idea of these methods is to introduce backdoor or Trojan horse watermarking [17, 29, 30] to certify the ownership of DL models, and only legitimate users can extract the full watermark.

In recent years, information hiding about DNN has become a popular research issue [18, 27, 31–37]. Kandi et al. [6] proposed an innovative learning-based autoencoder convolutional neural network (CNN) for nonblind watermarking, which adds an additional dimension to the use of CNNs for secrecy and outperforms methods using traditional transformations in terms of both agnosticism and robustness. Hayes and Danezis [37] used adversarial training techniques to learn a steganographic algorithm for the discrimination task. However, the DNN model of information hiding is radically different from other models in that if the DNN model is tampered, it means that the model parameters are also modified, reducing the accuracy of the image watermark detected by the model.

The abovementioned methods focused on protecting the model copyright. Meanwhile, the current study considers not only model copyright but also model integrity. Thus, in this paper, we propose a novel multipurpose watermarking method for protecting the copyright and integrity of a steganographic autoencoder network.

The main contributions of this work are summarized as follows:

(I) A method to protect DNN models by using multiple watermark association mechanisms is proposed. This method verifies not only the copyright information of the DNN model but also its integrity and can locate model tampering parts.

(II) The proposed work can ensure the accuracy of the image watermark extracted by the model according to the correlation between the model and image watermarks.

(III) The information hiding model adopts the average pooling method. Therefore, the designed symmetrical modification mechanism can ensure that the parameter mean value of the modified layers in the model remains relatively stable, so it has minimal impact on the average pooling results and ensures the stability of the model output.

The rest of this paper is structured as follows. First, we briefly describe HiDDen model and embedding strategy in Section 2, and then we detail the proposed method in Section 3 and demonstrate extensive experiments and analysis in Section 4. Finally, we conclude this paper in Section 5.

2. Related Works

2.1. HiDDen Model. A robust DNN model for data hiding was designed [10]. This approach generates visually indistinguishable watermarked images using an encoder given the input information and cover image. A decoder is also used to recover the input information from the encoded image. This model is robust against dropout, crop-out, cropping, Gaussian noise, and other image attacks, as shown in Figure 1.

The HiDDen model comprises the following four main components: an encoder E_θ, a decoder D_ϕ, a parameter-less noise layer N, and an adversarial discriminator A_γ. First, the watermark information W_1^{in} and the cover image I_{co} (size $C \times H \times W_1$) are fed into the encoder E_θ. The encoder E_θ then applies convolutions to the cover image to form a few intermediate representations and embeds the watermark information of length L in the encoder. After multiple convolutional layers process, the encoded image I_{en} is produced. Afterward, the noise layer N adds noise to the encoded image I_{en} to produce a noisy encoded image I'_{en}. Next, the noise-laden encoded image I'_{en} is fed to the decoder D_ϕ. This decoder D_ϕ then applies some convolutional layers to generate L feature channels in these intermediate representations. Global spatial average pooling and a fully

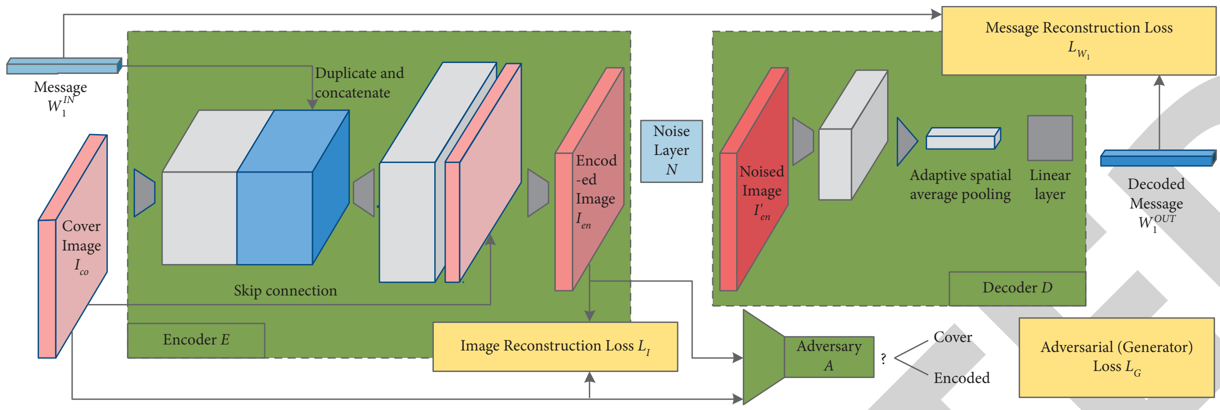


FIGURE 1: HiDDen flowchart.

connected layer are, respectively, applied to initialize a message vector M of the same size and then activated with a fully connected layer to decode the watermark W_1^{out}. The adversary A_γ is analogous to a decoder that serves to discriminate whether an image is an encoded image or a cover image and outputs a binary classification. The total loss function L_{Total} comprises L_{W_1}, L_G, and L_I and the associated loss function is defined below.

The loss L_I between the original image I_{co} and the encoded image I_{en} (image distortion loss) is defined by

$$L_I(I_{co}, I_{en}) = \frac{\|I_{co} - I_{en}\|_2^2}{CHW_1}. \quad (1)$$

The loss of the watermark information W_1^{in} and the decoded information W_1^{out} (watermark distortion loss) L_{W_1} is defined as

$$L_{W_1}\left(W_1^{\text{in}}, W_1^{\text{out}}\right) = \frac{\|W_1^{\text{in}} - W_1^{\text{out}}\|_2^2}{L}. \quad (2)$$

The L_G (adversarial loss) for the adversarial discriminator A_γ to detect whether an image is a watermarked image is defined as

$$L_G(I_{en}) = \log(1 - A(I_{en})), \quad (3)$$

where $A(I) \in [0, 1]$ is the probability of the watermarked image.

The classification loss L_A of the adversarial discriminator A_γ is defined as

$$L_A(I_{co}, I_{en}) = \log(1 - A(I_{co})) + \log(A(I_{en})). \quad (4)$$

In the original paper stochastic gradient descent on θ and ϕ is performed such that the total loss function L_{Total} is optimal in the following cases:

$$\mathbf{E}_{I_{co}, W_1}\left[L_{W_1}\left(W_1^{\text{in}}, W_1^{\text{out}}\right) + \lambda_I L_I(I_{co}, I_{en}) + \lambda_G L_G(I_{en})\right], \quad (5)$$

where λ_I and λ_G are regulators. Moreover, $\mathbf{E}_{I_{co}, W_1}[L_A(I_{co}, I_{en})]$ is minimized by training A_γ. At this point, the final decoded image is the watermarked image I_{em}.

2.2. Embedding Strategy for Model Watermarks with Modified Parameters. HiDDen trains robust coders and decoders using DNNs, but DNNs also require copyright protection. Thus, embedding watermark into a DNN is an excellent approach to prove its copyright ownership. The most typical method of watermark embedding is the parameter regularizer method adopted by Uchida et al. [14], by which a novel term is added into the initial cost function for the initial assignment. The cost function $E(\tau)$ with a regularizer is defined as

$$E(\tau) = E_O(\tau) + \lambda E_R(\tau), \quad (6)$$

where $E_O(\tau)$ is the original loss function, and $E_R(\tau)$ is the regularization term that imposes certain restrictions on parameter τ, and λ is an adjustable parameter.

Compared with the standard regularizer, the forced parameter w of this regularizer has a certain statistical deviation, which is used as the embedded watermark. This regularizer is called the embedded regularizer. Given a (mean) parameter vector $\tau \in R^m$ and an embedding key $X \in R^{T\times M}$, the watermark can be extracted only by using τ and X, and the threshold is set to 0. Specifically, the extraction of the j–th bit watermark is

$$w_j = s\left[\Sigma_i X_{ji}\tau(i)\right], \quad (7)$$

where $s(x)$ is a step function:

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{else}. \end{cases} \quad (8)$$

The flow of the algorithm is a binary classification problem with a single-layer perceptron. This means that it is straightforward to set up the loss function $E_R(\tau)$ for the embedding regularizer by using (binary) cross-entropy as a direct approach:

$$E_R(\tau) = -\sum_{j=1}^{T}\left[w_j \log_2\left(y_j\right) + \left(1 - w_j\right)\log_2\left(1 - y_j\right)\right], \quad (9)$$

where $y_j = \sigma(\Sigma_i X_{ji}\tau_i)$ and $\sigma(.)$ is the sigmoid function:

$$\sigma\left(\Sigma_i X_{ji}\tau_i\right) = \frac{1}{1 + \exp\left(-\Sigma_i X_{ji}\tau_i\right)}. \tag{10}$$

The loss function is applied to update τ instead of X. τ is the embedded target, and $X = (x_{ji})$ is the embedding key, $x_{ji} \sim N(0, 1)$. x_{ji} is embedded into each element about the parameter τ with random embedding weights.

3. Proposed Method

The flow diagram of the proposed algorithm is shown in Figure 2, and the details of three watermark embedding and extraction methods are described in this section. The HiDDen model introduced in [10] is selected as a carrier for the model watermarks W_2 and W_3 to authenticate the integrity of the DNN used for information hiding. The input for the HiDDen network is watermark W_1 and cover image I_{co}, and the output is the watermarked image. It includes three modules: an encoder E_θ, a decoder D_ϕ, and an adversarial discriminator A_γ, which can be trained jointly to be able to perform information hiding. Figure 1 shows the process of embedding the image watermark W_1 into the watermarked image I_{em} during the DNN training phase. In order to achieve multiple verifications of model integrity, in this work, we have modified the decoder module for the model by embedding additional model watermarks W_2 and W_3 to achieve multiple verifications of model integrity. This modification includes not only the model copyright information but also the image watermark information W_1 and the Hash values of the model parameters. In the model training phase, the original image is fed to this DNN model for training, and the final output is the watermarked image I_{em}. Blind detection of watermark information can be achieved in the watermark detection phase by extracting the output image watermark W_1 and the model watermark W_2. In addition, this work makes it possible to extract the model watermark W_3 and identify the tampering location of the DNN model when necessary. Model watermarks W_2 and W_3 will be embedded in the decoder of this model to protect its copyright.

The image watermark W_1 includes image copyright, comparison, and redundancy information. A certain region is divided into the other convolutional layers while selecting the fully connected layer in the DNN model to embed the model watermark W_2 to calculate the Hash value, which can initialize the model watermark W_3. The model watermark W_3 will be embedded in the redundancy parameters of the fully connected layer, which corresponds to the redundancy information of the image watermark W_1. The model watermark W_3 is extracted first to prove the integrity of the DNN model and locate the tampering location. The image watermark W_1 and the comparative model watermark W_2 can then be extracted and compared to determine the accuracy of the image watermark information. Thus, the copyright information of the image and model can be obtained.

3.1. Image Watermark W_1 for the Host Network. The HiDDen model proposed by Zhu et al. [10] is chosen in this work as a carrier. Compared with other models, the HiDDen model has the advantage of robustness to various attacks. The watermark embedded in the input image is referred to as the image watermark W_1 in this work. The image watermark W_1 comprises the following: image copyright information w_{11} and validation information w_{12}, $W_1 = \{w_{11}, w_{12}\}$.

3.2. Model Watermarks W_2 and W_3 in the Network. The ownership of the HiDDen model is further protected from copyright threats to enable cross-validation of watermarked information and identify the tampered location in the model. Suitable parameters in the convolutional and fully connected layers of HiDDen can be used in this work as carriers for model watermarks, thus achieving a small influence on the performance of the HiDDen model and an accurate location of the tampered parameter coefficients of the HiDDen model. To this end, model watermarks W_2 and W_3 are embedded in the DNN model in this work.

3.2.1. Model Watermarks W_2 and W_3 Generation. The structures of W_2 and W_3 are shown in Figure 3. Their specific compositions are as follows.

Composition of the model watermark W_2: model copyright information w_{21} and validation information w_{12}, $W_2 = \{w_{21}, w_{12}\}$.

Composition of the model watermark W_3: the chunked Hash values of all convolutional and fully connected layers constitute model watermark $W_3 = \{h_1, h_2, \ldots, h_{42}\}$.

3.2.2. Model Watermark W_2 Embedding Position. The proposed method generally embeds the model watermark on some layers of the network. For example, Uchida et al. [14] chose to embed the watermark on one of the intermediate layers of the network, while Feng et al. [36] embedded the watermark in multiple intermediate layers. Considering the suitable location for embedding the watermark, experiments revealed that embedding the watermark information in the middle layer closest to the output layer has the least impact on the model. Therefore, watermark information is embedded into the fully connected layer of the self-coding network.

The HiDDen model has the model parameters of the fully connected layer with size $L_1 \times L_1$. Thus, the model watermark W_2 with maximum capacity $L_1 \times L_1$ is denoted as $W_2 = \left\{W_2(1), W_2(2), \ldots, W_2(N_{W_2})\right\}$ $(\max(N_{W_2}) = L_1 \times L_1, N_{W_2} \leq L_1 \times L_1)$. The length of the model watermark W_2 was controlled to $L_2 \ll L_1 \times L_1$ considering the accuracy of the HiDDen model training (avoid excessive increase in watermark capacity). The effect of watermark capacity on the training accuracy of the HiDDen model is shown in Figure 4.

Each parameter in the HiDDen network is a 32-bit floating-point number, and the watermark is embedded in k decimal places. The imperceptibility of the algorithm

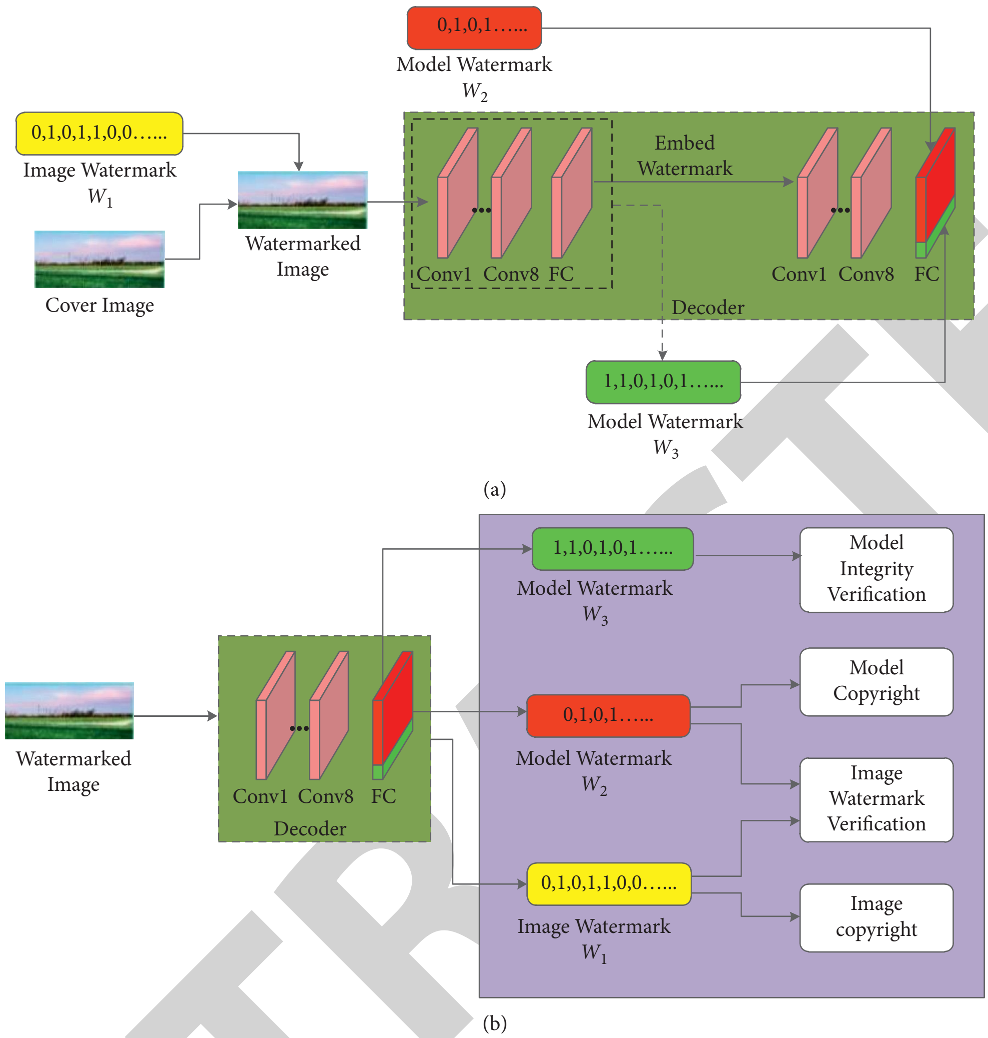


FIGURE 2: The proposed method framework. (a) Watermark embedding process and (b) watermark extraction process.

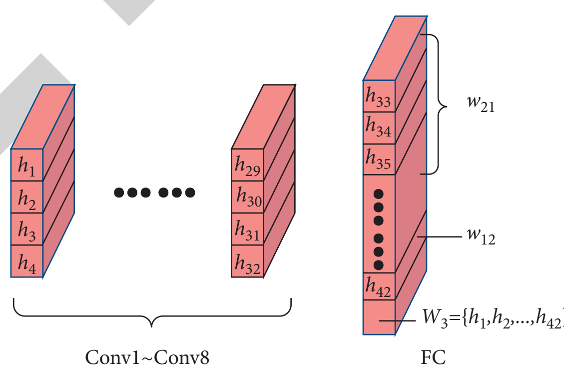


FIGURE 3: Composition of model watermarks $(W)_2$ and $(W)_3$.

increases with k but it is susceptible to truncation errors, weakening the robustness of the watermark extraction. Conversely, the robustness of the algorithm improves as k decreases. However, the accuracy of the HiDDen model is again affected, resulting in a decrease in model performance. Experimental verifications revealed that the performance of the HiDDen model is ideally balanced with the robustness of the watermarking algorithm with $k = 4$. The accuracy of the model for different k values is shown in Figure 5.

3.2.3. Model Watermark W_2 Embedding Strategy. Given a HiDDen model network with trained parameters, the mission of watermark embedding is defined as the

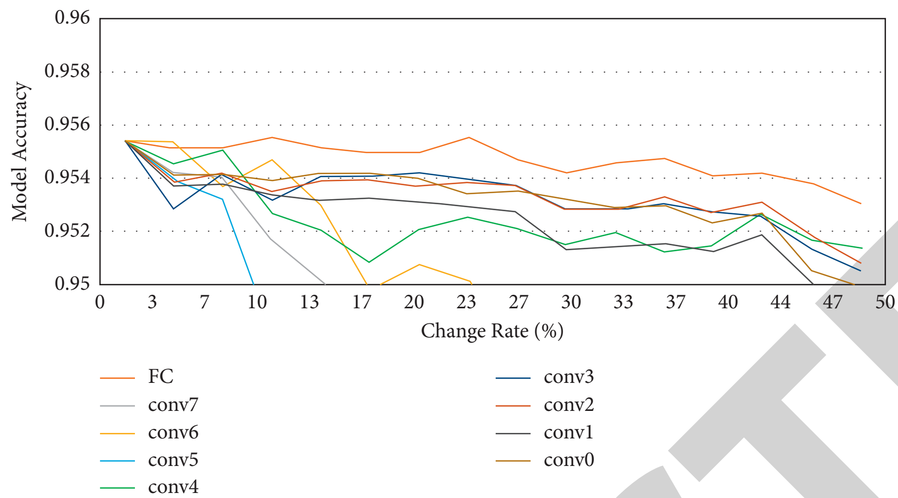


Figure 4: The effect of model watermark $(W)_2$ capacity on the training accuracy of the HiDDen model.

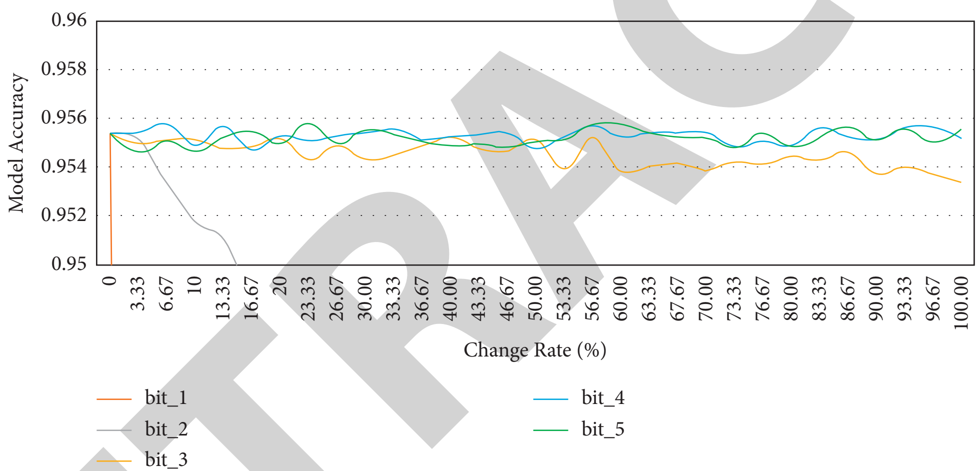


Figure 5: Model accuracy of different modification bits.

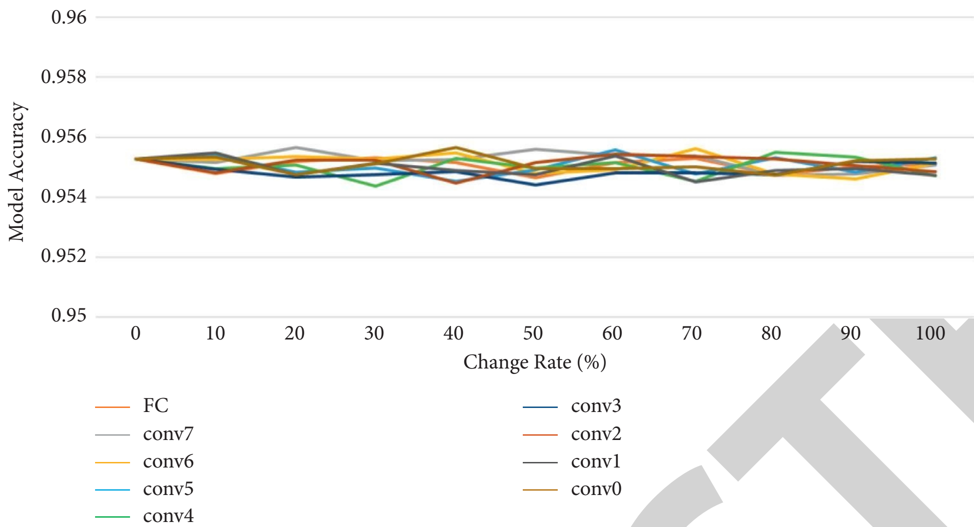
embedding of the model watermark W_2 to value D_k of the k–*th* decimal place of the fully connected layer model parameters.

Maintaining the model accuracy of decoders trained by neural networks is crucial when embedding watermarks. The HiDDen model performs average pooling on all convolutional layers, which reduces the impact on model accuracy if the mean value of the model parameters after watermark embedding is the same as that before embedding. Therefore, in this paper, we propose a symmetric watermark embedding strategy. The mean value is 4.5 assuming that numbers 0 to 9 fit the mean distribution at the k–position. The two states of the watermark are taken as (2, 7), which is a state pair as shown in Figure 6, to ensure that the mean value remains constant and the distance between the numbers is kept at a maximum. The specific embedding method is shown in

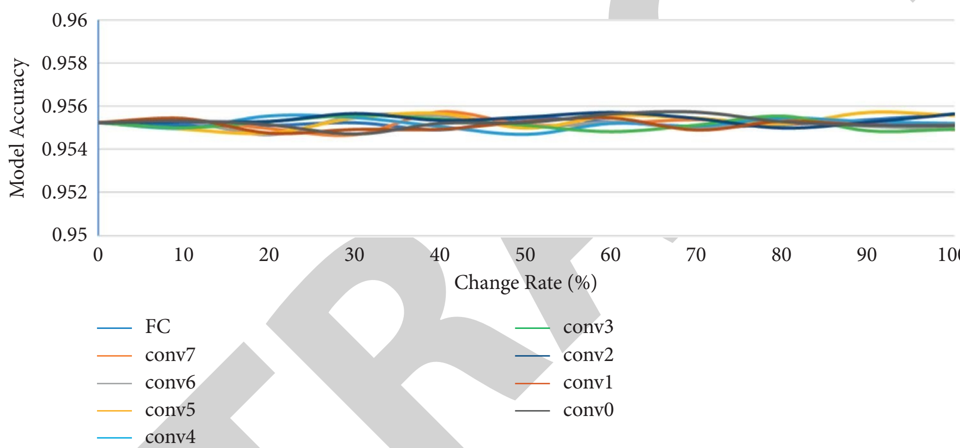
$$D_k = \begin{cases} 2, & \text{if } W_2(i) = 0, \\ 7, & \text{if } W_2(i) = 1. \end{cases} \tag{11}$$

The value of k in this paper is chosen within a median range. Therefore, the presences of the watermark neither affect the accuracy of the model nor are disturbed by quantization errors. At this point, the mean of the k–*th* bit is 4.5, which is equal to the mean of this bit of the model itself. The experimental data show no effects on the model accuracy when modified to lie in the fourth and subsequent decimal places. The watermark is embedded in all layers with minimal effect due to the slightly low bit count and for the convenience of extracting the model watermark W_2, which is embedded in the final fully connected layer in this work.

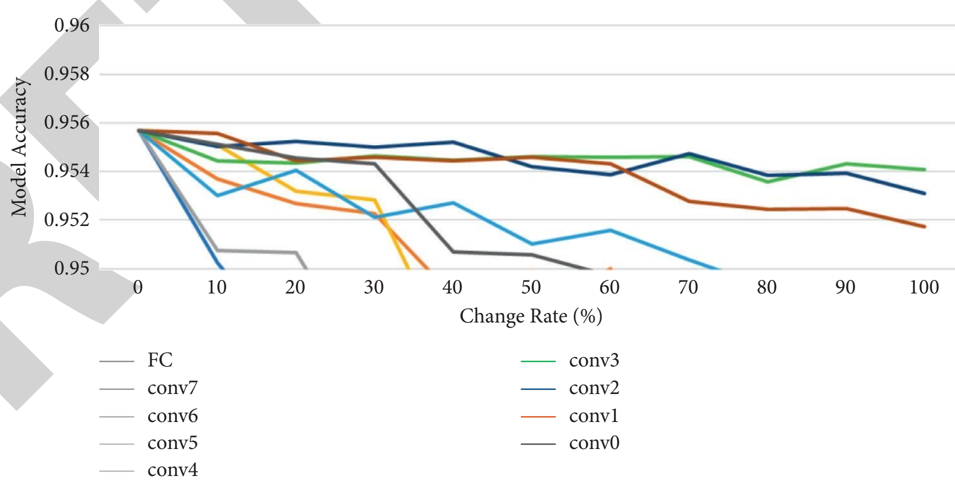
3.2.4. Model Watermark W_3 Embedding Position and Strategy. This work chunks the convolutional and fully connected layers of the HiDDen model to enable tampering localization. The small size of the block results in the large capacity of the model watermark W_3 and the high accuracy of the HiDDen model integrity certification. In practice, the different parameters can be freely chosen in accordance with



(a)

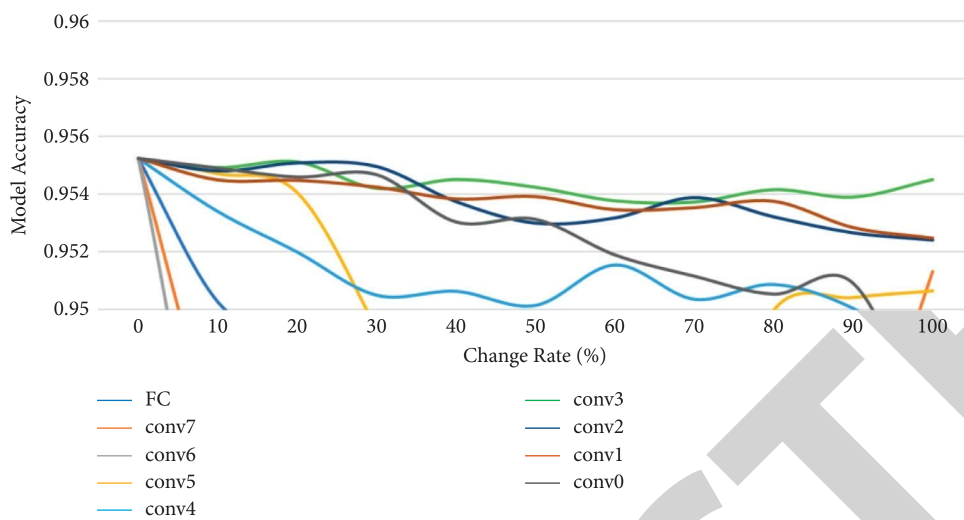


(b)



(c)

Figure 6: Continued.



(d)

Figure 6: Model accuracy of different embedding state pairs: (a) (2, 7) state pair, (b) (0, 9) state pair, (c) (3, 9) state pair, and (d) (7, 8) state pair.

the application needs, such as the capacity of the model watermark W_2.

Step 1. Calculate the Hash value of each block using the Hash function. These Hash values are known as the model watermark W_3, which is $W_3 = \{h_1, h_2, \ldots, h_{42}\}$,

Step 2. Write the Hash value of each chunk to the redundant bits of the fully connected layer. Therefore, the extracted Hash values of each block during HiDDen integrity verification can be compared with the model watermark W_3 for data integrity authentication.

Table 1 shows the corresponding experiments for different chunks and lists the effect of different numbers of chunks on model accuracy (the magnitude of change is 0.01).

3.3. Watermark Extraction. The three watermarks are extracted in reverse order of embedding, model watermark W_3, model watermark W_2, and then image watermark W_1.

3.3.1. Model Watermark W_3 Extracting. The model watermark W_3 is extracted according to the embedding rules; the coefficients corresponding to the eight convolutional layers and one fully connected layer in the decoder are found for chunking. The Hash value $\{h_1, h_2, \ldots, h_{42}\}$ of each block is then calculated and compared to the model watermark W_3 stored in the redundant bits of the fully connected layer. If they are equal, then no tampering will occur. Otherwise, the model block corresponding to h_i has been tampered.

3.3.2. Model Watermark W_2 Extraction. The watermark is embedded in the fully connected layer in the decoder. The watermark length of the model watermark W_2 is selected as the first L_2 model parameter in the fully connected layer in the decoder. The model watermark W_2 is then extracted in accordance with

$$W_2(i) = \begin{cases} -1, & \text{if } \mathrm{D}_k \in [04], \\ 1, & \text{if } \mathrm{D}_k \in [59]. \end{cases} \quad (12)$$

Model watermark W_2 and image watermark W_1 have the same validation information w_{12}. Thus, multiple validations of image watermark information and extraction of model copyright information can be achieved by comparing the detected model watermark W_2 and image watermark W_1.

3.3.3. Image Watermark W_1 Extraction. The watermarked image I_{em} is decoded into the HiDDen model, which first generates L feature channels using eight convolutional layers. A global spatially averaged pooling is then used to generate watermark vectors of the same size. The performance of the watermark decoder has been continuously improved after uninterrupted iterations of the coefficients in the fully connected layer [38]. Finally, the output image watermark W_1 is obtained through the fully connected layer.

4. Experiments

Experimental Evaluation: the hardware used for the experiments was a graphics card of NVIDIA GeForce RTX 3090/PCIe/SSE2, Intel® Core™ i9–10900X CPU @ 3.70 GHz × 20, and 62.5 GB memory. The standard Structure-Datasets applied for the experiments include coco-2014, coco-2017, and Boss.

4.1. Fidelity Assessment. In the proposed scheme, the coefficients of the embedded watermark are substantially smaller than the entire coefficients of the model. The watermark embedding takes an LSB-like approach, which has little impact on the model and hardly affects the model

Table 1: Model accuracy with different chunks.

Number of chunks per level	Modification rate of blocks (%)	Model accuracy (%)
2	4	92.47
	8	91.14
	12	90.20
	16	88.45
4	10	92.22
	20	90.89
	30	89.12
	40	87.55
5	10	92.46
	20	91.16
	30	90.29
	40	88.47
8	10	93.77
	20	92.01
	30	91.35
	40	90.84

Table 2: Fidelity evaluation of the three sets of our watermarking model.

Structure–Dataset	Class	Baseline accuracy (%)	Rate of watermarked weights (%)	Watermarked accuracy (%)
Coco-2014	All	96.02	15	95.51
	Cat	96.04	15	95.54
	Car	96.00	15	95.51
	Banana	95.94	15	95.51
	Person	95.88	15	95.43
Coco-2017	All	96.10	15	95.58
	Cat	96.16	15	95.52
	Car	96.00	15	95.47
	Banana	96.07	15	95.56
	Person	96.14	15	95.57
Boss	All	96.05	15	95.52

output accuracy. Taking standard Structure-Dataset coco-2014, coco-2017, and Boss as examples, the middle layer parameters of the model are approximately 223,812. The experiments show that the accuracy of the model does not diminish despite modification of 15% (33571) of parameters for watermark embedding as shown in Table 2. Figure 7 reveals the accuracy of different change rates. We refer to the HiDDen as the baseline accuracy and the accuracy of the watermarking model as the watermarking accuracy, and also separately for different kinds of images. The results indicate that the accuracy of the watermarked model is close to the baseline.

4.2. Image Quality. Only some layers of the decoder model in the HiDDen network are modified, and model watermarks W_2 and W_3 are embedded in the decoding layer of the self-coding network model. Thus, the quality of the output image is maintained despite the addition of the image watermark, as shown in Figure 8. Both the image watermarked and the final watermarked images of our proposed method have excellent visual quality compared with the original images.

4.3. Model Integrity Certification. The model watermark W_3 is extracted in accordance with the embedding rules of the watermark, and the Hash value of each block in each layer is also calculated and compared with the model watermark W_3. The corresponding blocks of the convolutional and fully connected layers corresponding to the Hash value h_i have been tampered with when the comparison of the Hash value h_i differs from that in the model watermark W_3. A digit after the decimal point is selected in the experiment to modify and embed the watermark (details are presented in subsection 3.2.4). Such a selection saves time and cost compared with that of Uchida et al. [14] and has advantages in watermark extraction accuracy. The test accuracy of the proposed watermarked model with different watermark capacities (in bits) is shown in Table 3.

4.4. Image Watermark Authentication. The model watermark W_2 and the image watermark W_1 have some mutual information between them. Thus, verification of the image watermark information and extraction of the model copyright information can be achieved by comparing the detected model watermark W_2 and the image watermark W_1.

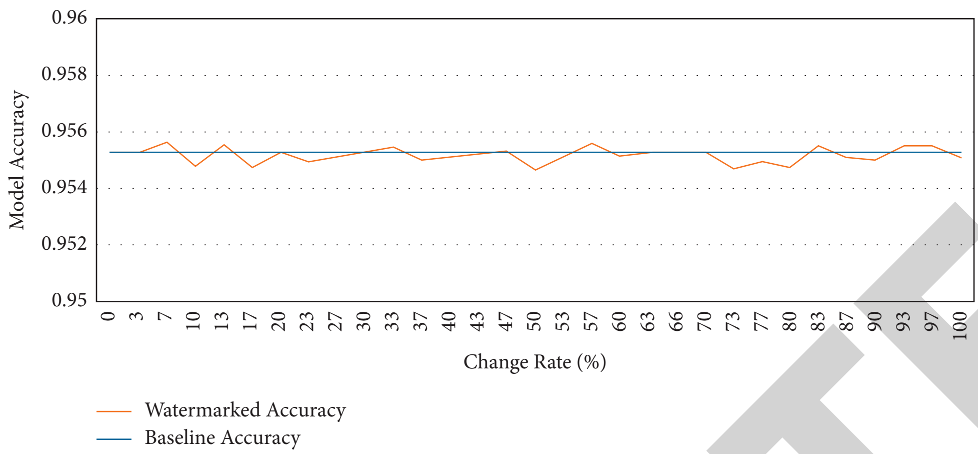


Figure 7: The comparison of watermarked accuracy and baseline accuracy.

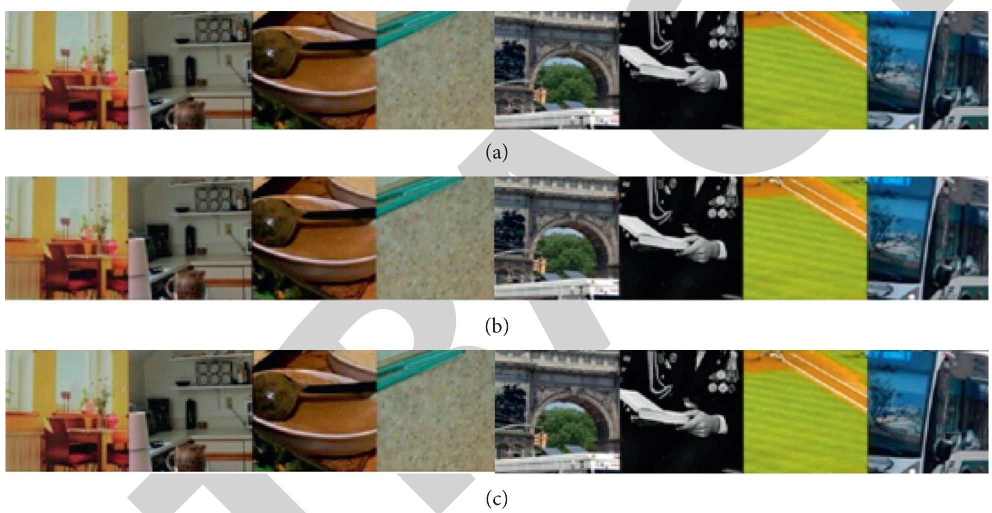


Figure 8: Comparison of image quality in three cases: (a) original images, (b) image watermarked images, and (c) final watermarked images.

Table 3: Test accuracy of watermarked model with different watermark capacities.

Embedded bits	Our watermarked model (%)	Uchida et al. (2017) (%)
256	95.5597	95.4542
512	95.5588	95.4563
1024	95.5543	95.4563
2048	95.5545	95.4588

5. Conclusion

In this paper, we propose an integrity authentication algorithm embedding multiple watermarks in the HiDDen model. These multiple watermarks include one image watermark and two model watermarks. The three watermarks are applied to protect the copyright information of the model and can pinpoint the exact location of model tampering. The fourth decimal place of the model parameters is modified to ensure the robustness and imperceptibility of the watermarking algorithm. The Hash values of all convolutional layers and fully connected layer are also used as one of the model watermarks for tampering location. Compared with previous algorithms, the proposed method achieves remarkable performance in various experiments considering fidelity, imperceptibility, model integrity authentication, and watermark authentication, rather than its practical value.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by Public Welfare Technology and Industry Project of Zhejiang Provincial Science Technology Department (no. LGG19F020016) and National Natural Science Foundation of China (no. 62172132).

References

[1] C. C. Lee, M. H. F. Rahiman, R. A. Rahim, and F. S. A. Saad, "A deep feedforward neural network model for image prediction," *Journal of Physics: Conference Series*, vol. 1878, no. 1, Article ID 012062, 2021.

[2] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, and Z. Zhu, "Deep speech 2: end–to–end speech recognition in English and Mandarin," in *Proceedings of the International Conference on Machine Learning*, pp. 173–182, New York, NY, USA, June 2016.

[3] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in *Proceedings of the. 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1715–1725, Berlin, Germany, August 2016.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing*, pp. 1097–1105, Mitpress, Cambridge, MA, USA, 2012.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large–scale image recognition," in *Proceedings of the International Conference on Learning Representations*, Sandiego, CA, USA, May 2015.

[6] H. Kandi, D. Mishra, and S. R. K. S. Gorthi, "Exploring the learning capabilities of convolutional neural networks for robust image watermarking," *Computers & Security*, vol. 65, pp. 247–268, 2017.

[7] S.-M. Mun, S.-H. Nam, H. Jang, D. Kim, and H.-K. Lee, "Finding robust domain from attacks: a learning framework for blind watermarking," *Neurocomputing*, vol. 337, pp. 191–202, 2019.

[8] K. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[9] G. Press, "Cleaning big data: most time–consuming, least enjoyable data science task, survey says. Forbes," 2019, https://www.forbes.com/sites/gilpress/2016/03/23/data–preparation–most–time–consuming–leastenjoyable–data–science–task–survey–says/#6d5fd596f637.

[10] J. Zhu, R. Kaplan, J. Johnson, and L. Fei, "HiDDeN: hiding data with deep networks," in *Proceedings of the European Conference on Computer Vision*, pp. 657–672, Munich, Germany, September 2018.

[11] H. Chen, C. Fu, B. D. Rouhani, J. Zhao, and F. Koushanfar, "DeepAttest: an end-to-end attestation framework for deep neural networks," in *Proceedings of the 46th International Symposium on Computer Architecture*, pp. 487–498, Phoenix, AZ, USA, June 2019.

[12] K. A. Zhang, A. Cuesta–Infante, L. Xu, and K. S.G. A. N. Veeramachaneni, "High capacity image steganography with GANs," 2019, https://arxiv.org/abs/1901.03892.

[13] M. Tancik, B. Mildenhall, and R. Ng, "StegaStamp: invisible hyperlinks in physical photographs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2117–2126, Long Beach, CA, USA, June 2019.

[14] Y. Uchida, Y. Nagai, S. Sakazawa, and S. Satoh, "Embedding watermarks into deep neural networks," in *Proceedings of the ACM International Conference on Multimedia Retrieval*, pp. 269–277, Bucharest, Romania, June 2017.

[15] M. Shafieinejad, N. Lukas, J. Wang, X. Li, and F. Kerschbaum, "On the robustness of backdoor–based watermarking in deep neural networks," 2019, https://arxiv.org/abs/1906.07745.

[16] T. Wang and F.. Kerschbaum, "Attacks on digital watermarks for deep neural networks," in *Proceedings of the 44th International Conference on Acoustics, Speech, and Signal Processing*, Brighton, United Kingdom, May 2019.

[17] Y. Liu, S. Ma, Y. Aafer et al., "Trojaning attack on neural networks," in *Proceedings of the 25nd Annual Network and Distributed System Security Symposium*, San Diego, CA, USA, February 2018.

[18] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38303–38314, 2018.

[19] Y. Nagai, Y. Uchida, S. Sakazawa, and S. i. Satoh, "Digital watermarking for deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 1, pp. 3–16, 2018.

[20] J. Wang, H. Wu, X. Zhang, and Y. Yao, "Watermarking in deep neural networks via error back–propagation," *Electronic Imaging*, vol. 22, no. 1-9, 2020.

[21] B. D. Rouhani, H. Chen, and F. Koushanfar, "DeepSigns: an end–to–end watermarking framework for ownership protection of deep neural networks," in *Proceedings of the Twenty–Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 485–497, April 2019.

[22] L. Fan, K. W. Ng, and C. S. Chan, "Rethinking deep neural network ownership verification: embedding passports to defeat ambiguity attacks," in *Proceedings of the Advanced Neural Information Processing System*, pp. 4716–4725, Montreal, Canada, November 2019.

[23] Y. Adi, C. Baum, M. Cisse, B. Pinkas, and J. Keshet, "Turning your weakness into a strength: watermarking deep neural networks by backdooring," in *Proceedings of the 27th USENIX Security Symposium (USENIX Security 18)*, pp. 1615–1631, USENIX Association, Baltimore, MD, USA, August 2018.

[24] J. Guo and M. Potkonjak, "Watermarking deep neural networks for embedded systems," in *Proceedings of the 2018 IEEE/ACM International Conference on Computer–Aided Design (ICCAD)*, pp. 1–8, San Diego, CA, USA, November 2018.

[25] J. Zhang, Z. Gu, J. Jang et al., "Protecting intellectual property of deep neural networks with watermarking," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, pp. 159–172, ACM, Incheon, Korea, June 2018.

[26] X. Zhao, H. Wu, and X. Zhang, "Watermarking graph neural networks by random graphs," 2020, https://arxiv.org/abs/2011.00512.

[27] R. Zhu, X. Zhang, M. Shi, and Z. Tang, "Secure neural network watermarking protocol against forging attack," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 1, p. 37, 2020.

[28] H. Wu, G. Liu, Y. Yao, and X. Zhang, "Watermarking neural networks with watermarked images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, 2020.