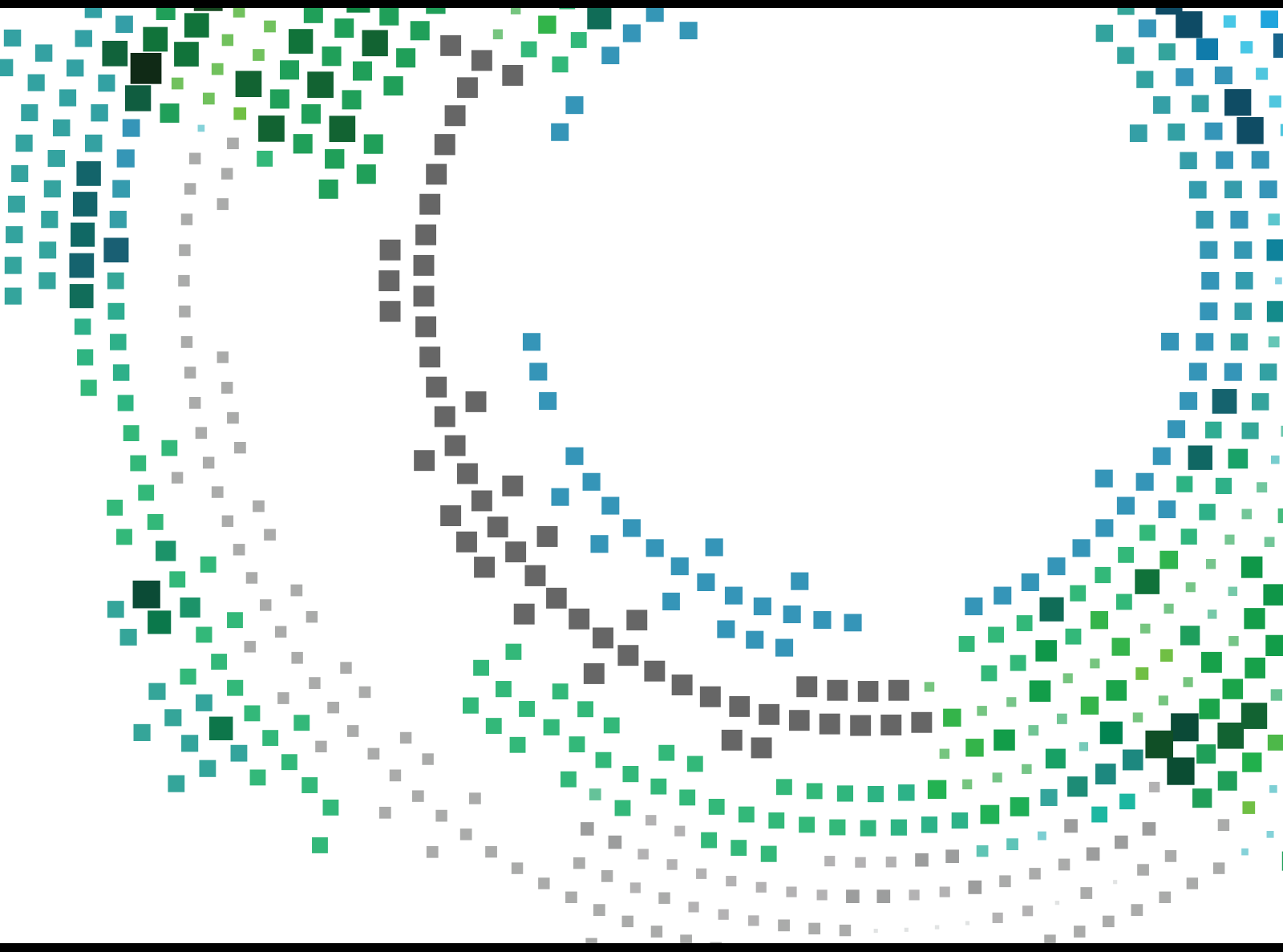
Mobile Information Systems

Machine Learning, Deep Learning and Optimization Techniques for Heterogeneous Sensor Information Integration 2022

Lead Guest Editor: Xingsi Xue
Guest Editors: Chin-Ling Chen, Miao Ye, and Pei-Wei Tsai



Machine Learning, Deep Learning and Optimization Techniques for Heterogeneous Sensor Information Integration 2022

Mobile Information Systems

Machine Learning, Deep Learning and Optimization Techniques for Heterogeneous Sensor Information Integration 2022

Lead Guest Editor: Xingsi Xue
Guest Editors: Chin-Ling Chen, Miao Ye, and Pei-Wei Tsai

Chief Editor

Academic Editors

Contents

Hindawi
Mobile Information Systems
Volume 2023, Article ID 8704278, 9 pages
https://doi.org/10.1155/2023/8704278

Research Article

Siamese Interaction and Fine-Tuning Representation of Chinese Semantic Matching Algorithm Based on RoBERTa-wwm-ext

Baohua Qiang,[1] **Guangyong Xi**,[1] **Yufeng Wang**,[2] **Xianyi Yang**,[1] **and Yuemeng Wang**[2]

[1]*Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin University of Electronic Technology, Guilin 541004, China*
[2]*Hebei Key Laboratory of Intelligent Information Perception and Processing, The 54th Research Institute of CETC, Shijiazhuang 050081, China*

Correspondence should be addressed to Xianyi Yang; xianyiyang65@126.com

Received 12 July 2022; Accepted 5 October 2022; Published 15 April 2023

Academic Editor: Xingsi Xue

Semantic matching research is the cornerstone of research in the fields of natural language similarity measurement and sensor ontology matching (OM). In the existing Chinese semantic matching methods, there are some shortcomings, such as the single dimension of semantic expression, the insufficient expression of context semantic relations, and the insufficient interaction of semantic information between different sentences. This paper proposes a Chinese semantic matching algorithm based on RoBERTa-wwm-ext with Siamese interaction and fine-tuning representation (RSIFR). The RSIFR model initializes the model with RoBERTa-wwm-ext as a vector of text. Firstly, a Siamese structure with embedded soft alignment attention mechanism and BiLSTM is constructed to realize the information interaction between two sentences. Secondly, LSTM-BiLSTM network structure is constructed to enhance the expression of semantic logic before and after sentences. Then, build a training model with fine-tuning mechanism. Fine tune the text's eigenvector parameters through label supervision. Finally, the fusion vectors of the sentence pairs are inserted into the MLP network layer, resulting in semantic matching results. RSIFR model starts from a variety of dimensions, strengthens the expression ability of vectors to text semantic relations, deeply mines the semantic similarities and differences between different sentences, and generally improves the Chinese semantic matching performance. Experiments on the public dataset LCQMC show that our model outperforms existing Chinese semantic matching models.

1. Introduction

Chinese semantic matching is to judge the semantic match between two different texts for them. The core of Chinese semantic matching task lies in mining the deep semantic information of text and exploring the semantic relationship between different texts. The research of text semantic matching can be applied to application areas such as intelligent question and answer, machine translation [1], natural language inference [2], WEB Sensor Ontology Matching (OM), and Entity Semantic Similarity Measurement [3].

The text feature vector extracted by using deep neural network technology can improve the vector's ability to characterize the text semantics, but it also lacks the representation of the semantic relationship between two sentences in the Chinese semantic matching task. Niu et al. combined Siamese network structure with deep learning techniques such as BiLSTM to effectively extract deep features of the text [4]. Yang and Zhang applied the attention mechanism in the Chinese semantic matching task to improve the representation of text by feature vectors [5]. The feature vector of the text extracted by the Chinese preprocessing model related to BERT [6] has stronger semantic expression ability, which effectively improves the performance of the Chinese semantic matching model, but lacks the semantic information interaction between different texts.

In order to enhance the interactivity of semantic information and improve the ability of vectors to represent the semantic matching relationship between texts, this paper

proposes a Chinese semantic matching algorithm based on RoBERTa-wwm-ext [7] with Siamese interaction and fine-tuned representation (RSIFR). RoBERTa-wwm-ext is used as the baseline model for Chinese preprocessing. For the preprocessing vectors, the Siamese interaction structure with embedded soft alignment attention mechanism and BiLSTM and the LSTM-BiLSTM network structure are built, which further enhance the representation ability of vectors. Then, a model for sentence pair classification fine-tuned based on RoBERTa-wwm-ext is built and pretrained, which is used to extract fine-tuned representation vectors of sentence pairs. The MLP structure is built for the final generated vector.

The main contributions of this paper are as follows:

(1) Input the two texts independently into RoBERTa-wwm-ext model and extract the Pooler_out layer vector of the model. For this vector, a Siamese interaction structure embedded with a soft-aligned attention mechanism and BiLSTM is built to enhance the semantic interaction between the two texts.

(2) Concatenate the two texts into a single-sentence text and input it into the RoBERTa-wwm-ext model and extract the Pooler_out layer vector of the model. An LSTM-BiLSTM network layer is built for this vector, which enhances the vector's expression of textual contextual semantic information.

(3) A training model that can fine-tune the initial vector of RoBERTa-wwm-ext is constructed, and a text vector fine-tuned by label supervision is constructed, which further improves the representation of the semantic relationship between texts by the vector.

1.1. Related Work. Semantic analysis is a fundamental task in various research fields such as text matching and ontology alignment (OA) [8]. The innovation of deep learning techniques provides a new technical support for semantic analysis tasks. Techniques such as RNN [9], CNN [10], and LSTM [11] are used to extract the features of text, which greatly improves the ability of feature vectors to characterize the semantic information of sentences. The updated iterations of the BERT series models have led to a breakthrough in the ability of vectors to express text semantics, providing research value in the field of Chinese semantic matching.

Deep neural network-based models are an important research direction in the field of Chinese semantic matching. Ranasinghe et al. used various combinations of GRU, Bi-LSTM, etc., in Siamese network structures to compare the representational power of various variants of the structure for text semantics [12]. Zhang et al. combined TF-IDF and Jaccard coefficients with CNN to improve the representation of vectors for sentence features, but lacked semantic links between different words [13]. Guo et al. analyzed the multiple semantic compositions of texts in terms of their frame structure and combined with the self-attention mechanism to enhance the representation of vectors for multiple semantic sentences, but lacked the representation of semantic relations between different texts [14]. Zhao et al. considered two granularities of words and characters in text and built a Siamese network structure containing BiLSTM, soft alignment attention mechanism to enhance the semantic interaction between text pairs [15].

The text feature vectors extracted by the BERT model exhibit good semantic representations [6], and various variant models incorporating BERT have emerged in the research area of text semantic matching. Peinelt et al. further enhanced the vector representation by incorporating sentence topic features based on the BERT model from the perspective of analyzing information about the topic elements of a sentence [16]. Viji and Revathy combined BERT with a BiLSTM-based Siamese structure, and they fed the vectors generated by BERT into the twin network for further hierarchical training to enhance the semantic representation of the vectors [17]. Srinarasi et al. used a combination of WordNet and BERT models to represent semantic features of text, further enhancing the representation of contextual semantic information within feature vector text [9]. Cui et al. proposed a whole-word masking approach to Chinese semantic training based on the BERT family of models for the structural properties of Chinese [7]. They constructed a series of Chinese pretraining models based on BERT, ALBERT [18], RoBERTa [19], etc., and applied the models to Chinese semantic matching tasks with relatively excellent performance.

In summary, the method based on deep neural network technology is characterized by its ability to effectively extract the contextual information of text semantics, but the ability to capture the semantic interactions between different texts is insufficient. The method based on pretrained models features effective representation of internal semantic relations of texts, but it lacks semantic interactivity between different texts. In this regard, this paper combines the RoBERTa-wwm-ext Chinese pretraining model with LSTM and BiLSTM, incorporates the SA-Attention (soft alignment attention mechanism), constructs Siamese interaction structures, and combines the RoBERTa-wwm-ext fine-tuned sentence pair classification model to improve the accuracy in Chinese semantic matching tasks.

2. Methodology

2.1. Model Framework. In this paper, we propose a Chinese semantic matching algorithm based on RSIFR. The model architecture is shown in Figure 1.

In Network Channel 1 (NC1), the two texts are independently connected to the RoBERTa-wwm-ext model to obtain the initial vector of the text. Then, a Siamese interaction structure with embedded SA-Attention_BiLSTM is built. The two initial vectors are crossed into the two Siamese channels and fused to produce the Siamese interaction type feature vector SiaVec.

In Network Channel 2 (NC2), we concatenate the two texts and feed the RoBERTa-wwm-ext model to extract the initial vector of the text. The initial vector is input to the LSTM-BiLSTM network layer to generate vector LBvec.

In Network Channel 3 (NC3), a sentence pair classification model based on RoBERTa-wwm-ext fine-tuning is

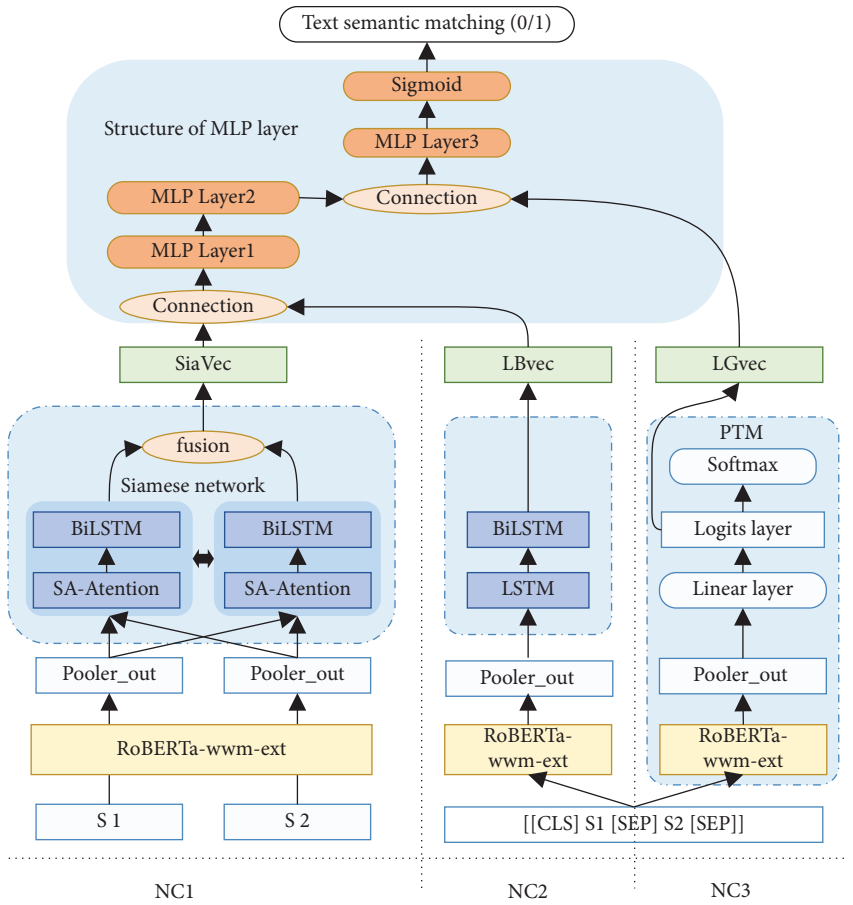


Figure 1: Model architecture diagram.

built, and a pretraining model PTM for sentence pair classification is generated for the dataset training, and the logit layer vector LGvec of the PTM is extracted.

At the MLP structure layer, vectors SiaVec and LBvec are connected and input to the first two fully connected layers of the MLP. Then, we concatenate the output vector with vector LGvec and feed the result into the last layer of the MLP, using sigmoid as the activation function to produce the final matching result of the sentence pair.

2.2. RoBERTa-wwm-ext Vectorization. RoBERTa-wwm-ext is a Chinese pretraining model, which adds whole-word masking (wwm) technology to the RoBERTa model and performs incremental training for large-scale Chinese data [7]. We use RoBERTa-wwm-ext as the baseline model of this model to initially extract the semantic features of Chinese texts to provide support for the downstream tasks of the model.

The two sentences S1 and S2 are input to the RoBERTaWE (RoBERTa-wwm-ext) model independently, and then the Pooler_out output layer vectors of the model are extracted separately. The formula is as follows:

$$\begin{aligned} S1\text{vec} &= [\text{RoBERTaWE}([S1])]_{Pooler\text{-}out}, \\ S2\text{vec} &= [\text{RoBERT}aWE([S2])]_{Pooler\text{-}out}, \end{aligned} \tag{1}$$

where S1vec and S2vec are the initial feature vectors of text S1 and S2, respectively.

2.3. Siamese Interaction Structure. The Chinese semantic matching task is to determine whether the meanings expressed by two different Chinese sentences are consistent. Siamese network is to input two matching Chinese texts into two Siamese subchannels independently, and two independent subchannels share training weights. The Siamese structure not only achieves the training independence of the two texts but also does not ignore the information interaction between the two texts. Within the two Siamese subchannels, the BiLSTM model is used to train the contextual semantic relationships of text, and the attention mechanism is used to enhance the interaction of text semantic information. The Siamese network structure based on SA-Attention and BiLSTM not only considers the learning of similar features between two sentences but also effectively exploits the heterogeneous information between two sentences, which enhances the performance of Chinese semantic matching tasks.

The Siamese interaction structure is shown in the NC1 channel in Figure 1. In the NC1 channel, the feature vectors of the two sentences are input into two Siamese subchannels, respectively. At the same time, after the feature vector of the sentence is processed by each layer of network structure, it is

connected with the vector before processing, so as to retain the original semantic features of the text and avoid the loss of information. Finally, the vectors computed by the two subchannels are fused to produce the final sentence pair vector representation. The network processing process is shown in Figure 2.

The vectors *S*1vec and *S*2vec are crossed input to the Siamese interaction structure and first processed by SA-Attention. The attention scoring function of vectors *S*1vec and *S*2vec is as follows:

$$Score\,(S1\text{vec}, S2\text{vec}) = S1\text{vec}^{T} \cdot S2\text{vec}, \tag{2}$$

where $Score\,(S1\text{vec}, S2\text{vec})$ is the attention scoring function, and then the attention distribution is calculated using the softmax function. The formula is as follows:

$$P = \frac{\exp\left(Score\,(S1\text{vec}, S2\text{vec})\right)}{\sum_{i=1}^{\dim} \exp\left(Score\,(S1\text{vec}, S2\text{vec})\right)_i}, \tag{3}$$

where P is the attention distribution function. We multiply P with the vectors *S*1vec and *S*2vec to calculate the corresponding weighted distribution. In order to avoid information loss, the initial vectors *S*1vec and *S*2vec are added to the calculation results. At the same time, the result of the addition is connected with *S*1vec and *S*2vec, respectively. The formula is as follows:

$$\begin{aligned} \text{Avec1} &= Concat\,[P \cdot S1\text{vec} + S1\text{vec},\ S1\text{vec}], \\ \text{Avec2} &= Concat\,[P \cdot S2\text{vec} + S2\text{vec},\ S2\text{vec}]. \end{aligned} \tag{4}$$

The vectors Avec1 and Avec2 are input to the BiLSTM, and the output vectors are concatenated with the vectors generated in the previous steps, respectively. The formula is as follows:

$$\begin{aligned} \text{SiaVec1} &= Concat\,[\text{BiLSTM}\,(\text{Avec1}), \text{Avec1}, S1\text{vec}], \\ \text{SiaVec2} &= Concat\,[\text{BiLSTM}\,(\text{Avec2}), \text{Avec2}, S2\text{vec}]. \end{aligned} \tag{5}$$

The vectors SiaVec1 and SiaVec2 are fused to produce the final vector representation SiaVec of the sentence pair. The formula is as follows:

$$\text{SiaVec} = [|\text{SiaVec1} - \text{SiaVec2}|, \text{SiaVec1} \odot \text{SiaVec2}], \tag{6}$$

where $\odot$ represents the multiplication of the corresponding terms of vectors SiaVec1 and SiaVec2. The vector SiaVec preserves the original semantic information of the text and enhances the semantic interaction between the two sentences.

2.4. LSTM-BiLSTM Network Structure. The LSTM model captures the semantic relationships between long-distance words in text very well, and it focuses on the forward encoding relationships in text sentences. The BiLSTM model focuses on both the positive and negative directions of the text, effectively expressing the context semantic relationship of the text. Chinese text semantics has a strong positive logical relationship. Based on this, we first use LSTM model to enhance the forward logical semantic representation of text semantics. Then, through the BiLSTM model, we pay more attention to the text semantic forward logic, but also learn the reverse semantic logical relationship of text. The LSTM-BiLSTM fusion model more effectively enhances the vector's contextual semantics for text.

The model is shown in Figure 1 for the NC2 channel. Firstly, two sentences S1 and S2 are connected to one sentence of text. The text *Sen* is input to the RoBERTaWE (RoBERTa-wwm-ext), and the output *P*vec of the Pooler_out layer is extracted as the initial vector representation of the text. The formula is as follows:

$$\begin{aligned} Sen &= [\,[\text{CLS}]S1\,[\text{SEP}]S2\,[\text{SEP}]\,], \\ P\text{vec} &= [\text{RoBERTaWE}\,(\,[\text{Sen}]\,)]_{Pooler\text{-}out.} \end{aligned} \tag{7}$$

Then, vector *P*vec is input to the LSTM layer, denoted as LSTM ([*P*vec]). To avoid losing information, first concatenate LSTM ([*P*vec]) with *P*vec, and then enter the BiLSTM layer to obtain the final vector LBvec. The formula is as follows:

$$\text{LBvec} = \text{BiLSTM}\,(Concat\,[\text{LSTM}\,(\,[P\text{vec}]), P\text{vec}]), \tag{8}$$

where LBvec is based on the RoBERTa-wwm-ext, which further enhances the semantic interaction within a single text and between two texts and enriches the representational information of the sentence pairs embedded in the vector.

2.5. Text Feature Representation Based on RoBERTa-wwm-ext Fine-Tuning. The feature vectors extracted directly from the RoBERTa-wwm-ext model ignore the influence of labels on the representation of text feature vectors. A RoBERTa-wwm-ext training model with fine-tuning mechanism is constructed, and the text's eigenvector parameters are adjusted through label supervision. The feature vector extracted by this structure covers the semantic association between text pairs, which improves the Chinese semantic matching performance.

The model is shown in Figure 1 for the NC3 channel. Firstly, the output vector *P*vec of the Pooler_out layer of RoBERTa-wwm-ext is input to the linear transformation layer. The formula is as follows:

$$\text{LWvec} = P\text{vec} \cdot W^{T} + Bias, \tag{9}$$

where W is the weight matrix of the vector *P*vec undergoing linear transformation and Bias is the bias of the function.

Then, the vector LWvec passes through the softmax activation layer, resulting in the final text pair matching result P_{LWvec}. The formula is as follows:

$$P_{LW\text{vec}} = \frac{\exp\left(\text{LWvec}\right)}{\sum_{i=1}^{\dim} \text{LWvec}_i}. \tag{10}$$

Supervised training is performed on the data to generate the sentence pair classification pretraining model PTM, and the logit output layer is extracted as a feature vector for the fine-tuned type of text pairs. The formula is as follows:

$$\text{LGvec} = [\text{PTM}\,(\,[Sen]\,)]_{Logits}. \tag{11}$$

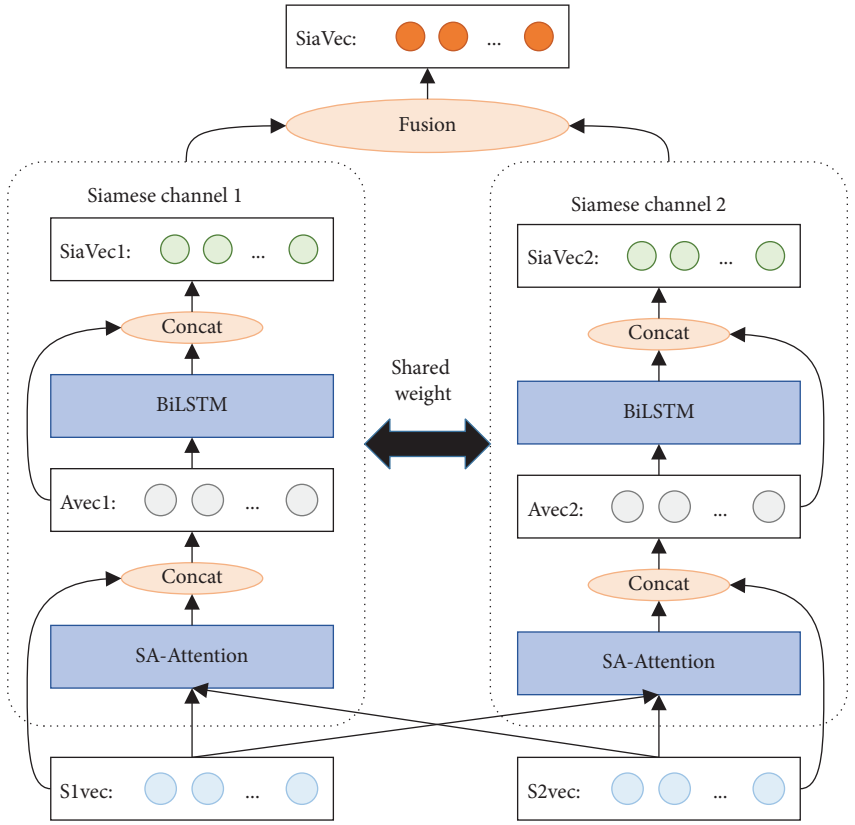


Figure 2: Siamese interaction structure.

The fine-tuned vector LGvec, which directly contains the semantic matching relationship between sentence pairs, plays a key role in the subsequent judgment of the matching degree of text pairs.

2.6. MLP Structure. After the analysis in the previous sections, vector LGvec contains the semantic matching information of sentence pairs, so vector LGvec is not involved in training in the first two fully connected layers of the MLP layer to avoid the loss of matching information.

Firstly, vectors SiaVec and LBvec are connected and participate in the training of the first two fully connected layers of MLP, and R_{L2} is the output of the fully connected layer. The formula is as follows:

$$\begin{aligned} R_{L1} &= \mathrm{MLP}_{layer1}(Concat[\mathrm{SiaVec}, \mathrm{LBvec}]), \\ R_{L2} &= \mathrm{MLP}_{layer2}([R_{L1}]). \end{aligned} \tag{12}$$

Then, the vectors R_{L2} and LGvec are connected to participate in the training of the fully connected layer of MLP layer 3, and finally, the final matching result is output by the sigmoid activation function. The formula is as follows:

$$\begin{cases} R_{L3} = \mathrm{MLP}_{layer3}(Concat[R_{L2}, \mathrm{LGvec}]), \\ R = \mathrm{MLP}_{Sigmoid}([R_{L3}])(R = 0\text{: } Mismatch, R = 1\text{: } Match), \end{cases} \tag{13}$$

where R is the matching result.

2.7. RSIFR Algorithm Implementation. The algorithm implementation of the model RSIFR is mainly divided into Chinese preprocessing and Chinese semantic matching classification training. The overall algorithm framework includes the entire processing process from the initial sentence pair input to the final semantic matching result, which more clearly shows the algorithm composition of RSIFR. The algorithm process is shown in Table 1.

3. Experiment

3.1. Dataset. The Chinese text contained in the public dataset LCQMC [20] covers a wide range of fields and is widely used in the research of Chinese semantic matching-related tasks. Therefore, in the experiments of this paper, we use LCQMC as the experimental dataset of this model. The size of the dataset is shown in Table 2.

The format of the LCQMC dataset is two Chinese texts corresponding to a 0/1 tag, with 0 indicating a semantic mismatch between the two texts and 1 indicating a match. The example of the dataset is shown in Table 3.

3.2. Ablation Experiments. The model RSIFR in this paper is designed with three core structures NC1, NC2, and NC3, as shown in Figure 1. To prove the validity and necessity of each structure in the model, the three modules NC1, NC2, and NC3 are eliminated on the basis of RSIFR model, and the evaluation index changes of RSIFR, $RSIFR^{\mathrm{NC1+NC2}}$, $RSIFR^{\mathrm{NC1+NC3}}$, and $RSIFR^{\mathrm{NC1+NC3}}$ models are

Table 1: RSIFR algorithm processing process.

RSIFR algorithm
1: Input: Sentence pairs S1, S2
2: Output: Semantic matching results R
3: S1, S2 independent input RoBERTa-WM, extract the pool_out layer vector SP1, SP2
4: S1, S2 connect input RoBERTa-WM, extract the pool_out layer P
5: Sentence pair classification model pretraining
6: for each epochs
7: S1, S2 input RoBERTa-WM, extract the pool_out layer vector D
8: Logits = linear (D)
9: res = softmax (logits)
10: end for
11: Training generates sentence pair classification pretraining model PTM
12: S1, S2 input PTM, extract logits layer vector LG
13: for each epochs
14: SP1, SP2 input SA-Attention, resulting in P1, P2
15: P1, P2 input BiLSTM, resulting in L1, L2
16: Concat [SP1, P1, L1] => SV1
17: Concat [SP2, P2, L2] => SV2
18: Concat [|SV1-SV2|, SV1* SV2] => SV
19: P input LSTM, resulting in PL
20: Concat [P, PL] input BiLSTM, resulting in LB
21: SV, LB, LG input MLP
22: Concat [SV, LB] Input MLP layer 1, resulting in R1. dropout (0.2)
23: R1 Input MLP layer 2, resulting in R2. dropout (0.2)
24: Concat [R2, LG] Input MLP layer 3, resulting in R3. dropout (0.2)
25: R = sigmoid ($R3$)
26: end for

experimentally compared. The evaluation index of the model utilizes the ACC and F1 values, and the results of the experiment are shown in Table 4.

In Table 4, the evaluation indexes ACC and F1 of RSIFR are greater than the evaluation indexes of each ablated model. This proves the necessity of the simultaneous existence of the three modules in the RSIFR model, and all three modules contribute to the performance improvement of the RSIFR model.

3.3. Performance Comparison Based on Different Baseline Models. RSIFR model is a vector representation with RoBERTa-wwm-ext to initially extract the text. Cui et al. published a series of Chinese preprocessing models such as BERT-wwm at the same time [7]. We used different Chinese preprocessing models as the baseline model for Chinese preprocessing in this paper model to compare and verify the Chinese preprocessing model with the best performance. The experimental data are shown in Table 5.

The experimental results show that RoBERTa-wwm-ext is used as the baseline model for Chinese preprocessing, and the two evaluation indicators of ACC and F1 on the LCQMC dataset achieve the maximum value, and the performance is the best.

3.4. Comparison of Existing Models. We compare the performance of RSIFR with existing models and use ACC and F1 as the evaluation indicators of the model. The data comparison is shown in Table 6.

Table 2: Size of LCQMC dataset.

Dataset	Total	Positive	Negative
Training	238766	138574	100192
Validation	8802	4402	4400
Test	12500	6250	6250

The existing BERT-related Chinese semantic matching model is a model derived from an improved model based on BERT. BERT-wwm is added with the method of full-word masking based on the BERT. BERT-wwm-ext has been trained incrementally in Chinese on the basis of BERT-wwm [7]. RoBERTa and MacBERT [7] are two other improved models of BERT. The two models were separately trained incrementally, resulting in Chinese semantic pretraining models such as RoBERTa-wwm-ext, RoBERTa-wwm-ext-large, MacBERT-base, and MacBERT-large [7]. They are applied to Chinese semantic matching tasks and show relatively good performance.

Among other existing models, Lattice-CNN is used to extract text semantic information from the perspective of multigranularity of text [21]; BiMPM is used to jointly capture the contextual semantic of the text from both positive and negative directions [22]; ESIM is used to use attention mechanism on text sequences to achieve inference between text sequences [23]; CATsNET is used to capture nonlocal features of text by building a Siamese network of crossattention mechanism [24]; GMN is used to build a graph structure that effectively expresses multiple textual meanings and combined with BERT [25];

Table 3: Example of LCQMC dataset.

Type	Sentence pair	Label
Positive	S1: 怎么能让玻璃更干净 (How can we make the glass cleaner) S2: 如何把玻璃擦得又干净又亮 (How to wipe the glass clean and bright)	1
Negative	S1: 你是学生还是老师 (Are you a student or a teacher) S2: 你是徐老师学生吗 (Are you Mr. Xu's student)	0

Table 4: Results of ablation experiments.

Model	ACC	*F*1
$RSIFR^{NC1+NC2}$	83.62	83.80
$RSIFR^{NC1+NC3}$	88.98	89.35
$RSIFR^{NC2+NC3}$	88.20	88.85
$RSIFR^{NC1+NC2+NC3}$ (ours)	89.56	89.67

Table 5: Experimental results based on different baseline models.

Model	ACC	*F*1
Based on BERT-base	88.90	88.84
Based on BERT-wwm	87.79	88.35
Based on BERT-wwm-ext	87.61	87.84
Based on ALBERT-base	88.50	88.20
Based on RoBERTa-wwm-ext-large	86.92	87.78
Based on RoBERTa-wwm-ext (ours)	**89.56**	**89.67**

Table 6: Comparison of existing models.

BERT-related models	ACC	*F*1	Other existing models	ACC	*F*1
BERT	86.9	—	Lattice-CNN	82.1	82.4
BERT-wwm	87.0	—	BiMPM	83.3	84.9
BERT-wwm-ext	87.1	—	ESIM	82.6	84.5
RoBERTa-wwm-ext	86.4	—	CATsNET	83.15	—
RoBERTa-wwm-ext-large	87.0	—	GMN	84.6	86.0
MacBERT-base	87.0	—	COIN	86.2	87.0
MacBERT-large	87.6	—	PERT	87.2	—
GMN-BERT	87.3	88.0	ABOEN	84.89	85.91
StyleBERT	87.9	—			
RSIFR (ours)	**89.56**	**89.67**	**RSIFR (ours)**	**89.56**	**89.67**

StyleBERT is used to combine Chinese pinyin, strokes, and other dimensions to enrich Chinese representation [26]; COIN is used for semantic alignment of different text sequences by establishing a context-aware crossattention mechanism [27]; PERT is used to establish a training method for text position substitution by combining N-gram and whole-word masking methods [28]; and ABOEN is an attention-based semantic enhancement model, which is used to extract finer-grained semantic information [29].

The two evaluation indicators of ACC and F1 of the RSIFR algorithm model achieved the maximum value in the comparison model, and the data are shown in Table 6. Experimental data show that the performance of the RSIFR algorithm model on the data LCQMC is better than the existing Chinese semantic matching model.

Before the RSIFR model worked, we locally prestored the initial vectors of the text. Subsequent text semantic analysis and text-matching tasks of the model are based on locally stored text initial vectors. In the process of model training, the time cost of initial text eigenvector initialization is saved, the training efficiency and running performance of the model are improved, and the training cost of the model is reduced.

4. Conclusions

In this paper, we propose a Chinese semantic matching algorithm based on RoBERTa-wwm-ext with Siamese interaction and fine-tuned representation. We study and design the Siamese interaction structure, the LSTM-BiLSTM network structure, and the text feature representation

structure fine-tuned based on RoBERTa-wwm-ext. The three structures generate twin interaction vectors, fully connected vectors, and fine-tuned representation vectors of sentence pairs, respectively, and a specific MLP network structure is designed for the three vector representations to obtain the final semantic matching result. The RSIFR Chinese semantic matching algorithm proposed in this paper starts from multiple dimensions of Chinese text, not only considers the contextual semantic relationship within a single text but also considers the semantic heterogeneous relationship between different texts, effectively strengthening the semantic interaction between different sentences It enhances the representation ability of vectors for textual context semantics. We show through experiments that the model proposed in this paper outperforms existing Chinese semantic matching algorithms on the public dataset LCQMC.

Data Availability

The data supporting the results of this study are public datasets. The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Guangxi under Grant 2019GXNSFDA185006; in part by the Development Foundation of the 54th Research Institute of China Electronics Technology Group Corporation under Grant SKX212010053; in part by the Development Fund Project of Hebei Key Laboratory of Intelligent Information Perception and Processing under Grant SXX22138X002; in part by the Guilin Science and Technology Development Program under Grants 20190211-17, 20210104-1; and in part by the Innovation Project of GUET Graduate Education under Grants 2022YCXS061.

References

[1] K. Seki, "Cross-lingual text similarity exploiting neural machine translation models," *Journal of Information Science*, vol. 47, no. 3, pp. 404–418, 2021.

[2] C. Liu, S. Jiang, and H. Yu, "Multi-turn inference matching network for natural language inference," in *Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 131–143, Springer, Guilin, China, August, 2018.

[3] X. Xue and C. Jiang, "Matching sensor ontologies with multi-context similarity measure and parallel compact differential evolution algorithm," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24570–24578, 2021.

[4] X. Niu, W. Zheng, and Y. Xiao, "Short Text Similarity Computation Method Based on Feature Expansion and Siamese Network," in *Proceedings of the 2021 4th International Conference on Data Science and Information Technology*, pp. 268–272, Springer, Singapore, July, 2021.

[5] Y. Yang and C. Zhang, "Attention-based multi-level network for text matching with feature fusion," in *Proceedings of the 2021 4th International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 1–7, Rochester, NY, USA, December, 2021.

[6] J. Devlin, M. W. Chang, and K. Lee, "Bert: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the NAACL-HTL*, pp. 4171–4186, Stroudsburg, PA, USA, 2019.

[7] Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang, "Pre-training with whole word masking for Chinese BERT," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3504–3514, 2021.

[8] X. Xue, "Complex Ontology Alignment for Autonomous Systems via the Compact Co-evolutionary Brain Storm Optimization Algorithm," *ISA Transactions*, vol. 99, p. 1, 2022.

[9] S. Srinarasi, R. Ram, and S. Raghavendra, "A Combination of Enhanced WordNet and BERT for Semantic Textual Similarity," in *Proceedings of the 2021 2nd International Conference on Control, Robotics and Intelligent System*, pp. 191–198, Rochester, NY, USA, October, 2021.

[10] B. Hu, Z. Lu, and H. Li, "Convolutional neural network architectures for matching natural language sentences," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[11] J. Mueller and A. Thyagarajan, "Siamese recurrent architectures for learning sentence similarity," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, Phoenix, AZ, USA, November, 2016.

[12] T. Ranasinghe, C. Orăsan, and R. Mitkov, "Semantic textual similarity with Siamese neural networks," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pp. 1004–1011, Varna, Bulgaria, September, 2019.

[13] P. Zhang, X. Huang, Y. Wang, C. Jiang, S. He, and H. Wang, "Semantic similarity computing model based on multi model fine-grained nonlinear fusion," *IEEE Access*, vol. 9, pp. 8433–8443, 2021.

[14] S. Guo, Y. Guan, R. Li, X. Li, and H. Tan, "Frame-based multi-level semantics representation for text matching," *Knowledge-Based Systems*, vol. 232, Article ID 107454, 2021.

[15] P. Zhao, W. Lu, and Y. Li, "Chinese semantic matching with multi-granularity alignment and feature fusion," in *Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Shenzhen, China, July, 2021.

[16] N. Peinelt, D. Nguyen, and M. Liakata, "tBERT: topic models and BERT joining forces for semantic similarity detection," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7047–7055, Red Hook, NY, USA, January, 2020.

[17] D. Viji and S. Revathy, "A hybrid approach of weighted fine-tuned bert extraction with deep siamese bi-lstm model for semantic text similarity identification," *Multimedia Tools and Applications*, vol. 81, pp. 1–27, 2022.

[18] Z. Lan, M. Chen, and S. Goodman, "Albert: a lite bert for self-supervised learning of language representations," in *Proceedings of the ICLR*, Ababa, Ethiopia, April 2020.

[19] Y. Liu, M. Ott, and N. Goyal, "Roberta: a robustly optimized bert pretraining approach," 2019, https://arxiv.org/abs/1907.11692.

[20] X. Liu, Q. Chen, and C. Deng, "Lcqmc: a large-scale Chinese question matching corpus," in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1952–1962, Stroudsburg, PA, USA, August 2018.

[21] Y. Lai, Y. Feng, and X. Yu, "Lattice Cnns for matching based chinese question answering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 6634–6641, Phoenix, AZ, USA, July, 2019.

[22] Z. Wang, W. Hamza, and R. Florian, "Bilateral multi-perspective matching for natural language sentences," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 4144–4150, Macao, China, February, 2019.

[23] Q. Chen, X. Zhu, and Z. Ling, "Enhanced LSTM for natural language inference," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 1657–1668, Stroudsburg, PA, USA, July 2017.

[24] Z. Wang, X. Zhang, and Y. Tan, "Chinese sentences similarity via cross-attention based siamese network," 2021, https://arxiv.org/abs/2104.08787.

[25] L. Chen, Y. Zhao, and B. Lyu, "Neural graph matching networks for Chinese short text matching," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6152–6158, Rochester, NY, USA, January, 2020.

[26] C. Lv, H. Zhang, and X. Du, "StyleBERT: chinese pre-training by font style information," 2022, https://arxiv.org/abs/2202.09955.

[27] Z. Hu, Z. Fu, and Y. Yin, "Context-aware interaction network for question matching," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3846–3853, Red Hook, NY, USA, April, 2021.

[28] Y. Cui, Z. Yang, and T. Liu, "PERT: pre-training bert with permuted language model," 2022, https://arxiv.org/abs/2203.06906.

[29] H. Zhang, H. Zhang, and X. Lu, "Attention-based overall enhance network for Chinese semantic textual similarity measure," *Journal of Applied Science and Engineering*, vol. 25, no. 2, pp. 287–295, 2021.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 3850246, 8 pages
https://doi.org/10.1155/2022/3850246

Research Article

Privacy-Enhanced Data Fusion for Federated Learning Empowered Internet of Things

Qingxin Lin,[1] Kuai Xu,[2] Yikun Huang,[2] Feng Yu,[3,4] and Xiaoding Wang[3,4]

[1]*Fuzhou University Zhicheng College, Fuzhou 350001, Fujian, China*
[2]*Concord University College Fujian Normal University, Fuzhou 350117, Fujian, China*
[3]*College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, Fujian, China*
[4]*Engineering Research Center of Cyber Security and Education Informatization, Fujian Province University, Fuzhou 350117, Fujian, China*

Correspondence should be addressed to Yikun Huang; fjnuhyk@163.com and Xiaoding Wang; wangdin1982@fjnu.edu.cn

Received 19 July 2022; Revised 1 September 2022; Accepted 15 September 2022; Published 3 October 2022

Academic Editor: Miao Ye

IoT sensors have already penetrated into extremely broad fields such as industrial production, smart home, environmental protection, medical diagnosis, and bioengineering. Although efficient data fusion helps improve the quality of intelligent services provided by the Internet of things, because the perceived data carry the sensitive information of the perceived object, the data fusion process is prone to the risk of privacy leakage. To this end, in this paper, we proposed a privacy-enhanced federated learning data fusion strategy. This strategy adds Gaussian noise at different stages of federated learning to achieve privacy protection in the data fusion process. Experimental results show that this strategy provides better privacy protection while achieving high-precision IoT data fusion.

1. Introduction

The Internet of things, also known as a sensor network, connects any item to the Internet through information sensing equipment such as radio frequency identification, infrared sensors, global positioning systems, and laser scanners for information exchange and communication, to achieve intelligent identification, positioning, tracking, monitoring, and management. The large-scale deployment and application of various sensors is an indispensable basic condition for the Internet of things. For example, different applications need to deploy different sensors, covering smart industry, smart security, smart home, smart transportation, smart medical care, etc. It can be seen that IoT sensor technology plays an important role in economic development and promotes social progress.

At present, most of the intelligent services provided by the Internet of things need to outsource user data to service providers for analysis and processing, which may easily lead to the leakage of sensitive information [1]. With the improvement of user privacy protection awareness and the promulgation of relevant laws and regulations, data analysis services based on traditional machine learning can no longer meet users' privacy protection needs. Although existing cryptographic technology can solve some privacy leakage problems, both symmetric encryption systems and asymmetric encryption systems have the risk of key leakage, and high-cost encryption chips cannot be popularized to terminal devices. In order to solve the problem of user data privacy leakage in related service scenarios, Google proposed federated learning technology [2, 3].

Federated learning is a distributed machine learning technology with privacy-preserving properties that can generate secure, accurate, and robust data models without analyzing the real data of users. In the intelligent service scenario based on federated learning, the service provider convenes different participants to provide data models by publishing federated learning tasks and aggregates all data models through the aggregation server to generate a reliable global model to provide related services. The reliability of the

global model is crucial, and a reliable global model can provide secure and stable services for IoT applications. It is worth noting that the use of cryptography technology can achieve data privacy protection. However, the use of cryptography often requires a trusted third party to generate a key for data encryption, which is difficult to achieve in the Internet of things. Compared with cryptography, federated learning does not require a trusted third party and is easy to deploy. Even if the federated learning server is not trusted, the privacy protection of data can be achieved by adding noise to the model, so federated learning has more advantages in data sharing.

The problem solved in this paper is formally described as follows, that is, how to realize data sharing under the premise of privacy protection. Based on the above analysis, a data fusion architecture is first proposed based on federated learning, as shown in Figure 1. The architecture consists of three layers: perception layer, data fusion layer, and intelligent service layer. Among them, the perception layer obtains perception data and sends the data to the data fusion layer through various sensors such as wearable sensors, vehicle-mounted sensors, surveillance cameras, and industrial sensors. In the data fusion layer, each federated learning (FL) data fusion center is responsible for the intelligent fusion processing of perception data and provides the fusion data to the intelligent service layer. This layer provides technical support for various intelligent services of the Internet of things such as intelligent transportation, smart grid, intelligent manufacturing, and intelligent logistics. All in all, the perception layer provides the necessary data to the intelligent service layer through the data fusion layer, and the intelligent service layer sends feedback information to it, hoping to improve the quality of the intelligent service.

According to this architecture, we consider using federated learning techniques to achieve privacy-enhanced data fusion. Furthermore, we combine differential privacy techniques with different stages of federated learning to further improve privacy protection during data fusion. The main contributions of this paper are as follows:

(1) To achieve privacy-preserving data fusion, we propose a privacy-enhanced federated learning data fusion strategy. This strategy not only adds differential privacy noise in the local model training process but also adds differential privacy noise in the federated training process, at the cost of a certain model accuracy, and the differential privacy protection of the local model and that of the global model are achieved simultaneously.

(2) Experimental results show that this strategy provides better privacy protection while achieving high-precision IoT data fusion.

The rest of this paper is organized as follows: Related work is described in the Related Work section. The system model is given in the System Model section. The specific implementation of the proposed strategy is elaborated in the Implementation Details of the Proposed Strategy section. Performance evaluations are given in the Performance Evaluation section. The Conclusions section concludes this paper.

2. Related Work

Federated learning for data fusion is an effective means for IoT to provide intelligent services, and the reliability of the federated learning global model determines the quality of services. More and more scholars at home and abroad have carried out research on how to ensure the reliability of the federated learning global model under different needs and have produced many excellent research results.

For federated learning task publishers, the reliability of the global model is the focus of attention. Researchers further ensure global model reliability by detecting anomalous models in the models to be aggregated. Cao et al. [4] mapped the local models into a graph through the Euclidean distance between local models and selected the local model for aggregation by solving the maximum clique problem in the graph, realizing the detection of anomalous models in federated learning. Zhao et al. [5] generated a dataset for auditing the local model through the trained generative adversarial network, and the prediction and evaluation results of the local model in the dataset were used as the criterion for judging whether it was an abnormal model, so as to realize the detection of abnormal models. Zhao et al. [6] proposed a proxy-based anomaly model detection mechanism, selecting participants with relatively stable performance in federated learning to perform anomaly model detection. Tolpegin et al. [7] extracted abnormal model features by performing dimensionality reduction and principal component analysis on the local model and realized abnormal model detection in the process. Liu et al. [8] proposed a federated learning scheme PEFL to mitigate poisoning attacks under privacy enhancement. In [9], an asynchronous update paradigm for real-time identification of client network parameters was proposed. This paradigm adopted a linear fusion method based on sequential filtering, considered communication delay, and asynchronously fused the parameters of the federation center. Then, a client real-time identification method based on linear filtering was established to obtain new label samples at unequal intervals, and the client was expected to have better performance.

For federated learning participants, the biggest demand is that federated learning can protect the private data of their training from being leaked. Since privacy and model reliability cannot be taken into account at the same time, existing research work mainly seeks a balance between the two, that is, reducing the loss of global model reliability while meeting the needs of participants for privacy protection. In [10], a privacy-preserving federated learning scheme, LDP-Fed, was proposed, which allows federated learning participants to protect the privacy of the model through personalized local differential privacy technology to prevent the leakage of deep information in the local model. Hu et al. [11] introduced differential privacy technology in federated learning and used the uncertainty brought by heterogeneity

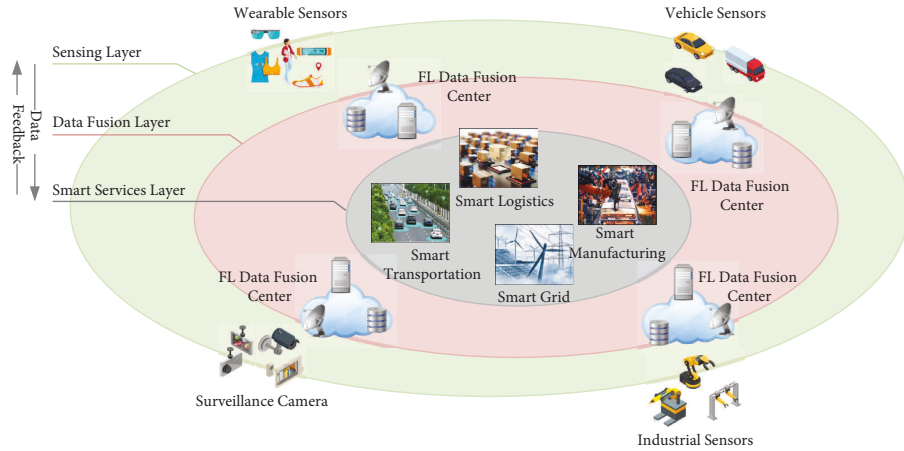


Figure 1: Federated learning-based data fusion architecture.

of IoT devices to perform differential privacy on the model to reduce the risk of privacy leakage. In [12], differential privacy technology and self-normalization technology were introduced in federated learning and the differential privacy noise layer and SELU security training layer were added during model training to realize the privacy protection of uploaded models. In [13], a blockchain-based federated learning training strategy was proposed, which uses differential privacy and homomorphic encryption technology to ensure the privacy of participants in the process of local model transmission and aggregation. Zhang and Luo [14] generated training data for local training by training a GAN model and proposed a new loss function, which enabled the generated training data to have indistinguishable visual features from the original data and protect the privacy of the training data of the participants. Liu et al. [15] used the sparse features of the feature map in the network model to represent the data that participants used for local training and realized the privacy protection of the real data. Xu et al. [16] proposed an efficient and privacy-preserving vertical-federated learning framework FedV, which implemented a two-stage noninteractive secure-federated aggregation method by introducing functional encryption and realized the privacy protection of real data of participants. In [17], Lin et al. proposed a secure joint learning mechanism based on variational autoencoders to resist inference attacks, in which participants reconstructed the original data through variational autoencoders, and trained local models on this basis to protect data privacy. In [18], a heterogeneous model fusion-federated learning mechanism was proposed, in which each node trained learning models of different scales according to its own computing power. After the parameter server received the training gradient of each node, it used the repetition matrix to correct the received gradient, then updated the corresponding region of the global model according to the mapping matrix, and finally assigned the compressed model to the corresponding node.

The above literature provides a large number of excellent algorithms for data fusion in the Internet of things. However, how to combine the privacy protection of local data with the privacy protection of the federated learning process to further strengthen the privacy protection of federated learning is still a problem worthy of research.

3. System Model

To achieve privacy-preserving data fusion in IoT, we need to consider the following three entities:

(i) Sensor (data provider): the sensor aggregates sensed data to the data fusion center through wireless or wired transmission.

(ii) Federated learning (FL)-based data fusion center: the data fusion center uses local sensor data for model training so that the local model can carry the information of local data. In addition, differential privacy noise needs to be added in the local model training process to achieve differential privacy protection of the local model.

(iii) Federated learning server: this server aggregates the local models of each data aggregation center to form a global model and adds differential privacy noise in the process to further improve the differential privacy protection capability of the global model.

3.1. Security Model. The privacy breach scenarios we consider are as follows. First, IoT smart service providers may be interested and commercialize private information about objects whose sensors are collecting data, thereby exposing their privacy, and federated learning can help reduce that risk. However, the aggregation server of federated learning may also be curious about the privacy of the perceived object, so there is also the risk of privacy leakage during the federated learning process. Differential privacy protection in local model training helps mitigate this risk. In addition, malicious attackers try to obtain the private information of perceptual objects from the global model through inference attacks. Adding differential privacy protection to the global model can effectively resist such attacks.

Input: Initial model parameter θ received from the FL fusion server, learning rate η, local sensor dataset D, gradient clipping C, privacy budget ϵ, sensitivity s, and Gaussian noise to be added satisfying $(\epsilon, \sigma) - DP$
Output: Final model parameter $\tilde{\theta}$
(1) **for** $t \in T$ **do**
(2) Calculate the gradient g_i for each batch $D_i \in D$
(3) Clip the gradient by $g_i = g_i/\max(1, |g_i|/C)$ and calculate average gradient $g = 1/|D| \sum g_i$
(4) Perform gradient descent by $\theta \leftarrow \theta - \eta g$
(5) Add Gaussian noise by $\tilde{\theta} = \theta + N(\sigma^2)$, where $\sigma^2 = 2s^2 \log(1.25/\sigma)/\epsilon^2$
(6) **end for**

Algorithm 1: Local data fusion with differential privacy protection.

Input: Local model parameter θ^k, privacy budget ϵ, sensitivity s, and Gaussian noise to be added satisfying $(\epsilon, \sigma) - DP$
Output: Global model parameter θ to be released
(1) **for** each FL-based data fusion center **do**
(2) Calculate the weighted average model by $\tilde{\theta} = \sum_{k=1}^{K} n_k/n\theta^k$
(3) Add Gaussian noise by $\theta = \tilde{\theta} + N(\sigma^2)$, where $\sigma^2 = 2s^2 \log(1.25/\sigma)/\epsilon^2$
(4) **end for**

Algorithm 2: Global data fusion with differential privacy protection.

4. Implementation Details of the Proposed Strategy

The data fusion strategy proposed in this paper is mainly composed of two modules, namely, the local data fusion module with differential privacy protection and the global data fusion module with differential privacy protection. The difference between these two modules is to add differential privacy noise to different stages of the federated learning process, thereby resisting privacy leak attacks on different objects.

4.1. Local Data Fusion with Differential Privacy Protection. Local data fusion is achieved by training a deep neural network model on local sensor data. In a deep neural network, by deploying multiple neurons at multiple levels and adjusting the connection weights between neurons by means of layer-by-layer training, the original feature data can undergo multiple nonlinear transformations. The fitting of any limited given input and output data finally obtains stable features for subsequent problem analysis. In the deep neural network algorithm, in order to evaluate the difference between the predicted value of the proposed neural network and the actual value, it is represented by a loss function L, and the mean square error loss function is used in this paper, i.e., $L(\theta, x) = 1/n \sum_{i=1}^{n} (y_i - x_i)^2$, where θ is the weight coefficient of the neural network to be trained, x represents the target value, y represents the predicted value output, and the subscript i represents the sample label. The purpose of deep neural network algorithm training is to minimize the loss function L. For complex neural networks, minimizing the loss function L is usually performed by stochastic gradient descent. That is, we randomly select training samples in batches during each iteration and calculate the partial derivative of the loss function L, denoted by $g = 1/|D| \sum_{x \in D} \nabla_\theta L(\theta, x)$, where D denotes the batches of samples, and then update the weight coefficient θ along the negative gradient direction towards the local minimum.

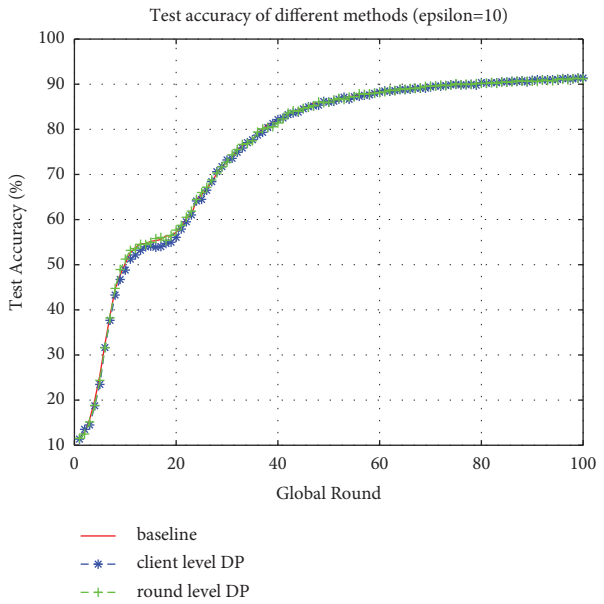
We adopt the differential privacy stochastic gradient algorithm whose objective function minimizes the loss function L by continuously training and adjusting the weight coefficients θ. The basic idea is as follows: in each iteration process, we first calculate the gradient of randomly generated batch samples $\nabla L(\theta, x_i)$, and gradient clipping is performed based on the L_2 norm of the computed gradient values. Considering the privacy protection of the sample data, the clipped gradient is updated with the mean value of the sum of the gradient and random noise based on the additional Gaussian noise method [19]. That is, by adding Gaussian noise with $\sigma^2 = 2s^2 \log(1.25/\sigma)/\epsilon^2$, the (ϵ, σ) differential privacy is achieved. Then, the weight coefficient θ of the next iteration is obtained. The implementation details of the local data fusion with differential privacy protection are summarized in Algorithm 1.

4.2. Global Data Fusion with Differential Privacy Protection. After the local model data fusion model is trained, the model will be sent to the federated learning server for aggregation. During training of the local model, we add Gaussian noise to resist privacy leakage attacks that may be launched by curious federated learning servers and IoT smart service providers. However, for inference attacks that malicious attackers may launch on the model, we add Gaussian noise again during the model release process to further enhance the model's privacy protection capabilities. The execution process of global data fusion with privacy protection is summarized in Algorithm 2.

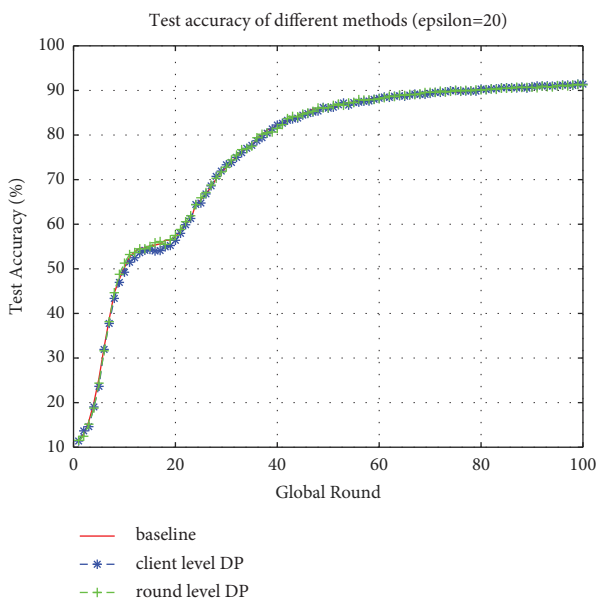
4.3. Security Analysis. The strategy proposed in this paper can resist privacy leakage attacks. First, in the training of the local model by the data fusion center, we add differential

Table 1: Parameter setup.

Hyperparameter	Value
Dp_δ	$1e-5$
Dp_ϵ	10, 20, 30
Epochs	100
Num_users	100
Frac	0.1
Local_ep	1
Local_bs	100
Learning rate	0.01
Lr_decay	0.995

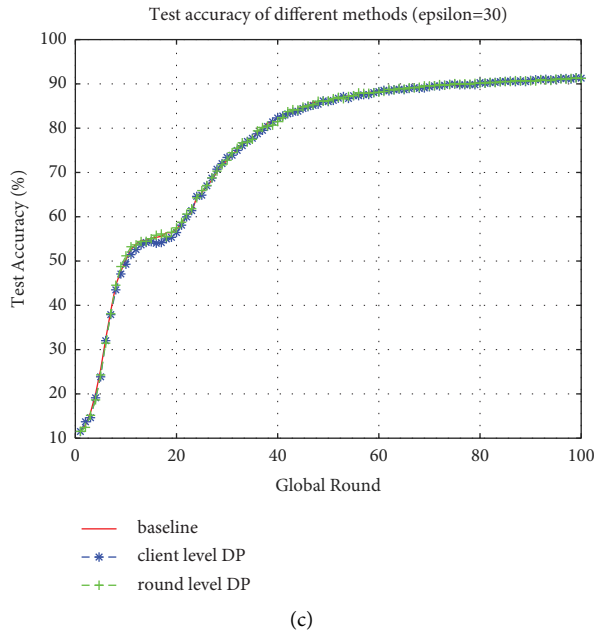


(a)



(b)

Figure 2: Continued.



(c)

Figure 2: Accuracy on the MNIST dataset with different differential privacy protections: (a) $\epsilon = 10$, (b) $\epsilon = 20$, and (c) $\epsilon = 30$.

privacy noise to the gradient of the model update so that the local model can get differential privacy protection after the local model training is completed. Second, after the local model is aggregated to the federated learning server, the server adds differential privacy noise to the aggregated model again and then distributes the noise model to each data fusion center so that the global model can get stronger differential privacy protection. We repeat the above process until federated learning converges. Due to the post-processing properties of differential privacy, the entire federated learning process has differential privacy protection. In addition, since the model is protected by differential privacy, it increases the possibility of attackers recovering user data through reasoning attacks and also increases the difficulty of attackers launching known plaintext attacks and ciphertext only attacks.

5. Performance Evaluation

5.1. Experimental Environment. The experiment is conducted to evaluate the performance of the proposed strategy on the computer equipped with i7 6.4GHZ processor, 32G memory, and win7 64-bit system. Federated learning is constructed through the Python-based deep learning framework (Tensorflow 2.2.0).

In the experiment, both the local model and the global model use CNN, which has 2 convolutional layers ($1 * 10$, kernelsize = 5; $10 * 20$, kernelsize = 5), dropout layers, and two fully connected layers ($320 * 50$; $50 * 10$). The datasets used in this experiment are the MNIST dataset and the Fashion- MNIST dataset. The MNIST dataset is a widely used handwritten digit recognition dataset, commonly used for performance evaluation of image classification algorithms in the field of computer vision. There are 10 digit classes in this dataset, from digit 0 to digit 9. The MNIST dataset contains 70,000 grayscale images with a resolution of $28 * 28$, of which 60,000 images are used for training the model and another 10,000 images are used for validation. The Fashion-MNIST dataset is an extended version of MNIST. The Fashion-MNIST dataset contains 70,000 grayscale images, including a training set of 60,000 images and a test set of 10,000 images. Each is a $28 * 28$ grayscale image, including different types of t-shirts, dresses, and boots. In the experiments, we fix other hyperparameters and adjust $\epsilon \in (10, 20, 30)$ for multiple experiments. The rest of parameters are given in Table 1.

In this experiment, we compare noise-added local model training, denoted by client level DP, noise-added global model training, denoted by round level DP, and the baseline strategy (FedAvg) [20] in terms of model accuracy.

6. Experimental Results

Figures 2 and 3 show the accuracy of our proposed strategy for local model training and global model training under different privacy budgets, i.e., $\epsilon \in (10, 20, 30)$. It can be observed from Figure 2 that the accuracies of three strategies all rise rapidly before 20 rounds, then slowly rise after that, and converge to the optimal accuracy when approaching 100 rounds, which is about 90%. In addition, local model training and global model training do not have lower accuracy than FedAvg under the same number of epochs due to the addition of Gaussian noise. Furthermore, under different privacy budgets, the accuracy of the local model and the global model is not much different. It can be observed from Figure 2 that the accuracy of three strategies increases rapidly before 30 rounds, then slowly increases, and converges to the optimal accuracy, which is about 70%,

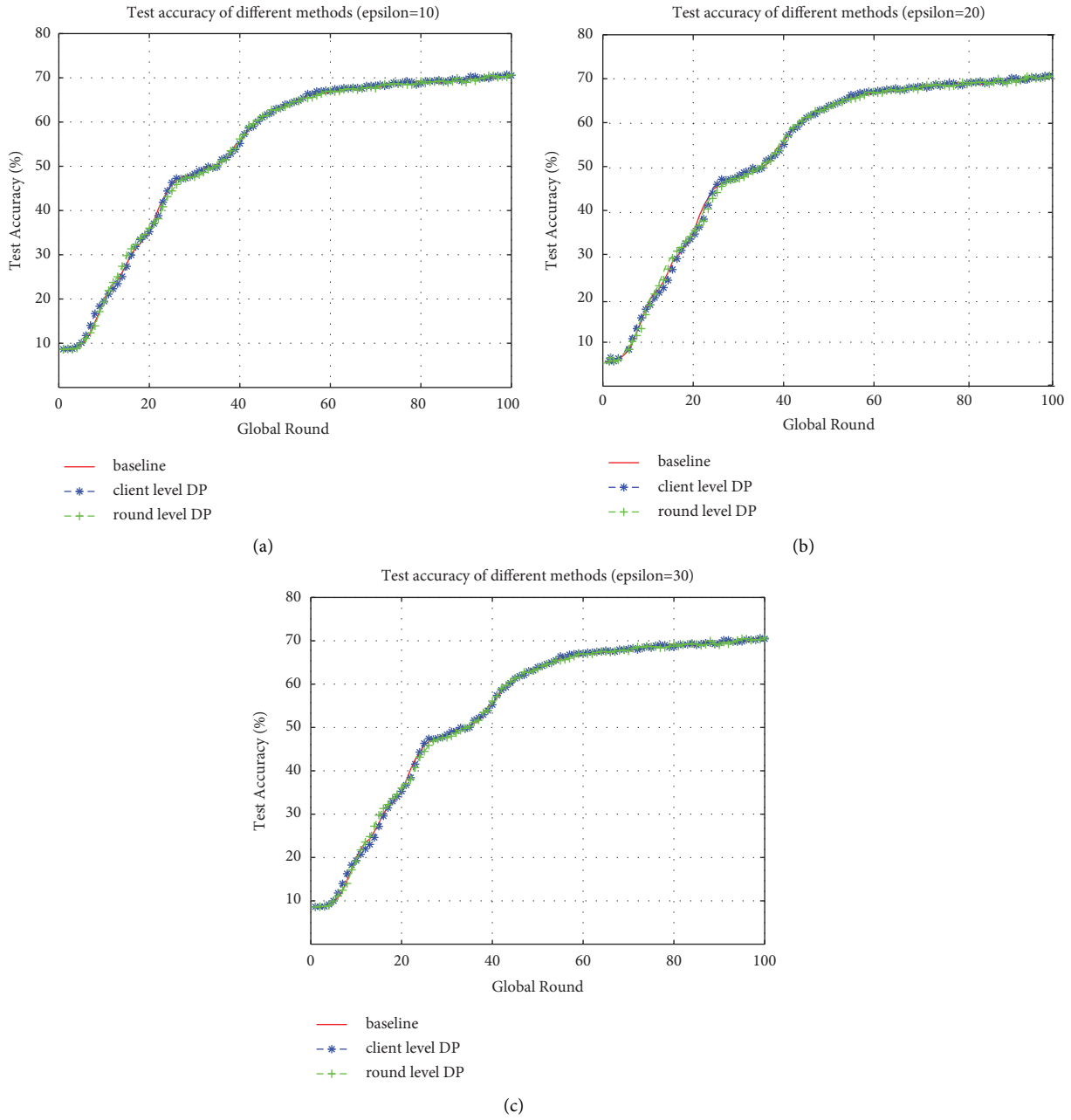


Figure 3: Accuracy on the Fashion-MNIST dataset with different differential privacy protections: (a) $\epsilon = 10$, (b) $\epsilon = 20$, and (c) $\epsilon = 30$.

when approaching 100 rounds. In addition, the accuracy of local model training and global model training under the same number of epochs is close to FedAvg, and under different privacy budgets, the accuracy of the local model and the global model is not much different.

7. Conclusions

The deep fusion of data collected by various sensors in the Internet of things is an urgent problem to be solved. In addition, in the process of data fusion, the privacy of objects collected by sensors may be leaked due to data fusion, which means the necessity of data fusion and privacy protection. To this end, we propose a privacy-enhanced federated learning data fusion strategy. This strategy not only adds differential privacy noise in the local model training process but also adds differential privacy noise in the federated training process, so as to realize the differential privacy protection of the local model and the differential privacy protection of the global model at the same time. Experimental results and theoretical analysis show that this strategy provides better privacy protection while achieving high-precision IoT data fusion. Considering that the addition of

noise will affect the accuracy of the model, our future research directions include how to reduce the impact of noise on the global model accuracy under different local models.

Data Availability

The Fashion-MNIST data used to support the findings of this study have been deposited in the repository, that is, https://github.com/zalandoresearch/fashion-mnist.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant nos. U1905211 and 61702103 and the Natural Science Foundation of Fujian Province under Grant nos. 2020J01167 and 2020J01169 and 2022J01644.

References

[1] M. Al-Rubaie and J. M. Chang, "Privacy-preserving machine learning: threats and solutions," *IEEE Security & Privacy*, vol. 17, no. 2, pp. 49–58, 2019.

[2] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: distributed machine learning for on-device intelligence," 2016, https://arxiv.org/abs/1610.02527.

[3] J. Konečnỳ, H. B. McMahan, X. Yu Felix, P. Richtárik, T. S. Ananda, and D. Bacon, "Federated learning: strategies for improving communication efficiency," 2016, https://arxiv.org/abs/1610.05492.

[4] Di Cao, S. Chang, Z. Lin, G. Liu, and D. Sun, "Understanding distributed poisoning attack in federated learning," in *Proceedings of the 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*, Tianjin, China, December 2019.

[5] Y. Zhao, J. Chen, J. Zhang, Di Wu, M. Blumenstein, and S. Yu, "Detecting and mitigating poisoning attacks in federated learning using generative adversarial networks," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 7, Article ID e5906, 2022.

[6] L. Zhao, S. Hu, Q. Wang et al., "Shielding collaborative learning: mitigating poisoning attacks through client-side detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, p. 1, 2020.

[7] T. Tolpegin, S. Truex, M. E. Gursoy, and L. Liu, "Data poisoning attacks against federated learning systems," *Computer Security - ESORICS 2020*, vol. 32, pp. 480–501, 2020.

[8] X. Liu, H. Li, G. Xu, Z. Chen, X. Huang, and R. Lu, "Privacy-enhanced federated learning against poisoning adversaries," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4574–4588, 2021.

[9] X. Ma, C. Wen, and T. Wen, "An asynchronous and real-time update paradigm of federated learning for fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8531–8540, 2021.

[10] S. Truex, L. Liu, Ka-Ho Chow, M. E. Gursoy, and W. Wei, "Ldp-fed: federated learning with local differential privacy," in *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, Heraklion, Greece, April 2020.

[11] R. Hu, Y. Guo, H. Li, Q. Pei, and Y. Gong, "Personalized federated learning with differential privacy," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9530–9539, 2020.

[12] O. Ibitoye, M. O. Shafiq, and A. Matrawy, "Differentially private self-normalizing neural networks for adversarial robustness in federated learning," *Computers & Security*, vol. 116, Article ID 102631, 2022.

[13] S. Kumar, S. Dutta, S. Chatturvedi, and M. P. S. Bhatia, "Strategies for enhancing training and privacy in blockchain enabled federated learning," in *Proceedings of the 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pp. 333–340, IEEE, New Delhi, India, September 2020.

[14] X. Zhang and X. Luo, "Exploiting Defenses against gan-based Feature Inference Attacks in Federated Learning," 2020, https://arxiv.org/abs/2004.12571.

[15] B. Liu, Y. Guo, and X. Chen, "Pfa: privacy-preserving federated adaptation for effective model personalization," *Proceedings of the Web Conference*, vol. 2021, Article ID 3449847, pp. 923–934, 2021.

[16] R. Xu, N. Baracaldo, Yi Zhou, A. Ali, J. Joshi, and H. Ludwig, "Fedv: privacy-preserving federated learning over vertically partitioned data," in *Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security*, Los Angeles , CA , USA, September 2021.

[17] H. Lin, W. Liu, and X. Wang, "A secure federated learning mechanism for data privacy protection," in *Proceedings of the 2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS)*, London, United Kingdom, December 2021.

[18] X. Lu, Y. Liao, C. Liu, P. Lio, and P. Hui, "Heterogeneous model fusion federated learning mechanism based on model mapping," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 6058–6068, 2022.

[19] C. Dwork, A. Roth, and C. Xiao, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2013.

[20] B. McMahan, E. Moore, D. Ramage, S. Hampson, and y A. Blaise Aguera, "Communication-efficient learning of deep networks from decentralized data," *Artificial intelligence and statistics*, vol. 20, pp. 1273–1282, 2017.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 8378187, 14 pages
https://doi.org/10.1155/2022/8378187

Research Article

Anomaly Detection in QAR Data Using VAE-LSTM with Multihead Self-Attention Mechanism

Chuitian Rong,[1] **Shuxin OuYang**,[1] **and Huabo Sun**[2]

[1]*School of Computer Science and Technology, Tiangong University, Tianjin 300384, China*
[2]*Institute of Aviation Safety, China Academy of Civil Aviation Science and Technology, Beijing 100028, China*

Correspondence should be addressed to Shuxin OuYang; 2031081023@tiangong.edu.cn

Received 22 June 2022; Accepted 8 September 2022; Published 30 September 2022

Academic Editor: Chin-Ling Chen

With the rapid development of the aviation industry, it is particularly important to ensure the safe flight of aircraft. How to find potential hazards in the process of aircraft flight has always been one of the important topics of civil aviation research. At present, the Quick Access Recorder (QAR) is the most widely used equipment to store the data recorded on aircraft. QAR data contain a lot of valuable and unexplored information, which records the true status of the aircraft in detail. Therefore, finding abnormal data from QAR data lays an important foundation for obtaining the cause of abnormality and providing a guarantee for flight. In this paper, in order to discover the abnormal information in the QAR data, we applied a VAE-LSTM model with a multihead self-attention mechanism. Compared to the VAE and LSTM models alone, our model performs much better in anomaly detection and prediction, detecting all types of anomalies. We conducted extensive experiments on real-world QAR data sets to prove the efficiency and accuracy of our proposed neural network model. The experimental results proved that our proposed model can outperform state-of-the-art models under different experimental settings.

1. Introduction

With the continuous growth of civil aviation passenger traffic, aviation safety has become a significant issue in the world. Safety is the prerequisite for the steady development of all industries and the basis for the survival of the transportation sector. Ensuring aviation safety has always been a major challenge for aviation activities [1].

Flight Operational Quality Assurance (FOQA) is an important scientific method of aviation safety management, which is used to monitor the data generated by aircraft. Over the last few decades, with improved sensing capabilities, there are different recorders that have been installed on aircraft to monitor the aircraft systems and flight crew performance. In these recorders, the Quick Access Recorder (QAR) is easier to install and configure compared to the Flight Data Recorder (FDR) and Cockpit Voice Recorder (CVR). QAR is a flash recorder for aircraft data acquisition systems and is also a key data source for airlines to evaluate flight quality and aircraft engine operations [2]. It covers most of the parameters of aircraft flight, including aircraft attitude parameters and engine-related data. By analysing a number of flight parameters recorded by QAR, the anomalies can be detected to avoid safety hazards and improve flight quality.

Nowadays, aircraft failure detection and early warning based on QAR data have become one of the important fields of civil aviation scientific research. However, there are many factors that can affect the quality of QAR data, such as working environment, signal transmission, data precision, and data decode computation Therefore, the original QAR data contain many anomalies and cannot be used directly without processing. In order to improve the quality of QAR data, it is necessary to perform anomaly detection on QAR data. Anomaly detection from QAR data is also one of the important strategies of FOQA. Finding anomalies from QAR data in time can prevent many unnecessary losses. To ensure the safe flight of the aircraft, it calls for an efficient and accurate anomaly detection method using advanced techniques.

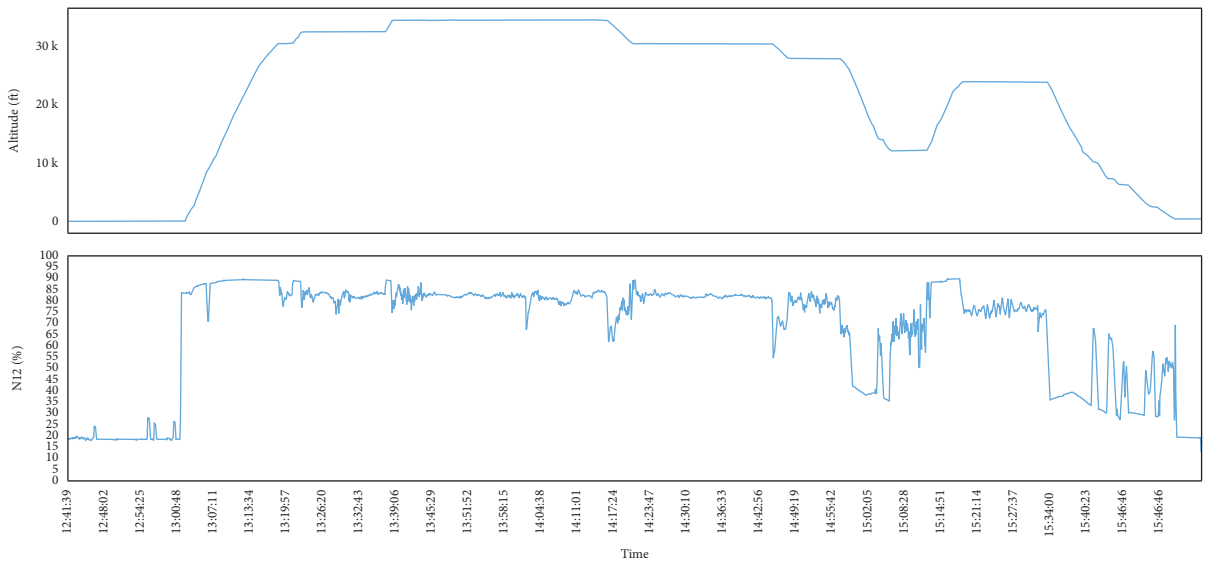


Figure 1: Example of QAR data.

QAR data records hundreds of parameters in each flight, including $N11$ and $N12$ engine thrust parameters, altitude, and vertical acceleration. It is one type of multivariate time series data. It also has the characteristics of a large amount of data, strong temporal structure, and regularity of change trends. Compared with classical time series data, QAR data have some peculiar features. The classical time series data are collected from stationary sensors. The QAR data are collected during the flight, which is divided into multiple flight phases. In different phases, the flight aircraft is working at different status and environments. Thus, the recorded parameters in QAR data present different distributions. Figure 1 takes the altitude parameters and the thrust parameters of the $N12$ engine in the QAR as an example. In actual flight, QAR data record the whole flight of an aircraft from take-off to landing, including *TAXI OUT, TAKE OFF, 2 SEGMENT, INI. CLIMB, CLIMB, CRUISE, DESCENT, APPROACH, FINAL, LANDING, and TAXI IN*, 11 phases in total, as shown in Figure 1.

In recent years, many scholars have proposed many methods for anomaly detection of time series, among which deep learning methods are popular increasingly among scholars. Due to the process of forecasting and anomaly detection with a large amount of time series data, it is unrealistic to label a large amount of data for training a model. Therefore, unsupervised anomaly detection is the preferred solution for most scholars. Anomalies can be divided into three types according to different manifestations, namely, point anomalies, collective anomalies, and background anomalies. Point anomalies are the easiest to find and can usually be marked by simple threshold or clustering methods. In contrast, collective and background anomalies are the most common in life and deserve more in-depth study. Dealing with background anomalies usually takes into account the relationship between adjacent data, and the use of models based on predictive methods is very effective for detecting such anomalies. For example, Ergen and Kozat [3] used algorithms based on long short-term memory (LSTM) neural networks to identify anomalies by calculating the difference between predicted and actual values. Collective anomalies are usually subsequences or anomalies in the entire sequence. The first step in detection is usually to divide the time series into equal-sized windows and treat the extracted subsequences as the entire sequence. For example, methods based on auto-encoder (AE) [4] and variational auto-encoder (VAE) [5], which utilize reconstruction differences for anomaly detection, have been shown to be effective.

The separate VAE model only considers the time dependence within the window and cannot analyze the information outside the window. This paper proposes a VAE-LSTM hybrid deep model based on a multihead self-attention mechanism, which integrates VAE and LSTM as a whole for unsupervised anomaly detection. Instead of directly inputting the raw data into the LSTM model like other methods, we pretrain the VAE model first and then use the low-dimensional feature vector generated by the encoder as the input of the LSTM model. Using the VAE model to effectively capture the contextual information in the window enables the LSTM model to learn longer correlations in time series. First, after pretraining the VAE model, the encoder is used to divide the QAR data into windows of a specific shape to extract the features of the recorded parameters in the QAR data, and the generated low-dimensional feature vector is used as the input of the LSTM model. Next, we use the LSTM model to train the data for the memory function of the time series. We also improve the LSTM model by incorporating a multihead self-attention mechanism, which is derived from the Transformer model [6]. Attention mechanisms are usually used in related fields such as text classification and text translation. In this paper, we apply it in time series anomaly detection. The self-attention mechanism can adjust the weight of the data, which is equivalent to a feature extraction of the data itself, and it is easier to capture

long-distance interdependent features. The multihead self-attention mechanism operates multiple self-attention mechanisms in parallel, reducing the amount of computation by reducing the dimension. Finally, the feature vector generated by the multihead self-attention mechanism module is reconstructed by the decoder for anomaly detection. We utilize a multihead self-attention mechanism for deep feature extraction. Because compared with traditional deep learning methods, it can explore hidden features without relying on complex neural network structures and have higher efficiency and performance than them. It makes it easier to capture long-distance interdependent features. We believe that combining the multihead self-attention mechanism with the LSTM model can better focus on the long-term dependencies of time series. In this way, our model can effectively detect both short-term anomalies and long-term anomalies.

In summary, the main contributions of this work are as follows:

(i) We first pretrain the VAE model and optimize the model by maximizing the ELBO loss for feature learning. We propose a novel anomaly detection model for QAR data.

(ii) We improve the LSTM model by incorporating a multihead self-attention mechanism to capture long-term correlations in QAR data. It is able to detect all types of exceptions.

(iii) In order to further improve the classification accuracy, we adopt a threshold selection method that maximizes the $F1$ metric, which effectively reduces false positives caused by improper threshold selection.

(iv) We conducted extensive experiments on the real QAR dataset to evaluate our model and compared it with other deep learning methods. Experiments show that our model has a significant improvement over other methods.

2. Related Works

QAR data are multivariate time series data with unique characteristics compared to classical time series data. Few works focus on the anomaly detection of QAR data. In this section, we first discuss existing state-of-the-art methods in the field of anomaly detection and analyze their strengths and weaknesses in order to justify our proposed method.

Anomaly detection of time series has always been a complex and challenging task in many disciplines and has been widely studied by many scholars. In anomaly detection, temporal continuity is important. Outliers are often those that are defined as unusual due to a lack of continuity in their short or long history. Therefore, anomalies in time series can be divided into two categories: short-term anomalies and long-term anomalies. Short-term anomalies occur when there are sudden changes in series values or short time intervals in the time series. The long-term anomaly is the entire time series or a subsequence that is identified as anomalous. In the past, the field of anomaly detection has generated a large amount of literature. We can roughly classify their proposed methods into three categories: statistics-based methods, classical machine learning-based methods, and deep learning-based methods [7].

2.1. Statistical Methods. The most common methods are autoregressive moving average (ARMA) and one of its generalizations, differential autoregressive moving average (ARIMA) [8]. They are one of the classic prediction-based anomaly detection models and are suitable for univariate time series. The ARIMA model uses previous data to fit a linear equation for prediction, describes the relationship between current and historical values, and uses its own historical data to predict new data. It requires that the sequence be stationary, and for nonstationary sequences, it needs to be stationary by difference. Ottosen and Kumar [9] used the ARIMA anomaly detection technique to detect short-term anomalies in low-cost air quality datasets by calculating prediction errors based on the absolute value of the residuals. However, the disadvantage of this method is that it can only predict phenomena related to the previous data, and the number of autoregressions and the parameters of prediction error need to be selected appropriately.

2.2. Machine Learning Methods. Common machine learning algorithms include clustering methods such as K-means clustering [10]. The K-means algorithm is the basic and most widely used partitioning algorithm in clustering methods. The sample data are clustered by the specified number of categories K, and the corresponding cluster centroids are used to detect anomalies in the monitoring data. Li et al. [11] proposed a cluster-based algorithm to detect excessive QAR events. It converts each flight data into a high-dimensional vector and uses the DBSCAN algorithm to cluster the matrix row vectors. The purpose is to identify exceptions without knowing the normative standard. Zhao et al. [12] proposed an algorithm based on a Gaussian mixture model (GMM) that incrementally updates the clusters according to the data instead of reclustering and adapts to the new data through an expectation-maximization algorithm to handle dynamically changing data in flight data. Zeng et al. [13] used a density-based DBSCAN clustering method to detect aircraft onboard and controller data that deviate from the normal range. Edward Smart et al. [14] proposed a two-stage approach based on a support vector machine (SVM) classifier to detect anomalies in the descent stage of a specific flight. The first stage quantifies anomalies at specific altitudes during the flight, and the second stage ranks all flights to identify the most likely anomalies. Although the above algorithms can detect abnormal flights from QAR data, they do not take into account the temporal patterns between the data and do not better explain why the abnormality occurs.

2.3. Deep Learning Methods. Compared with the above two methods, deep learning-based anomaly detection models can capture more complex hidden features and temporal

correlations in time series, so they have received extensive attention in recent years. Broadly speaking, they can be divided into two categories: predictive models and generative models. Predictive models detect anomalies based on the error of the prediction as an anomaly score. In particular, convolutional neural network (CNN), recurrent neural network (RNN), and an improved model based on it, long short-term memory network LSTM [15], have achieved remarkable results. They all have a powerful ability to learn from data. In addition, LSTM networks can model longer data; it has control structures (gates) to regulate stored memory and learn and capture normal behavior. When encountering data that deviate significantly from normal data, it predicts a large error to indicate anomalies. Hundman et al. [16] used an LSTM model for the prediction of spacecraft telemetry data and used dynamic thresholding of errors to identify anomalies. Khorram et al. [17] combined CNN and LSTM as a novel convolutional long-short-term memory recurrent neural network for fault detection, achieving high generalization accuracy and resistance to overfitting. However, such a separate prediction model is not only computationally expensive but also may lead to large deviations in the prediction results due to some uncertain factors. LSTM models are also very sensitive to the choice of parameters. As a result, many advanced generative models have emerged, including variational autoencoder (VAE) [18] and generative adversarial networks (GAN) [19]. At their core, they learn representations of normal patterns. Kishore et al. [20] proposed a deep autoencoder (DAE) applied to the raw time series data of multiple aircraft sensors and used the error of AE reconstruction to determine whether the data were abnormal. Combining convolutional neural network (CNN) with VAE, Memarzadeh et al. [21] developed a convolutional variational autoencoder (CVAE) applied to the data of abnormal commercial flight departures. Wang et al. [22] proposed a sequential parameter attention-based convolutional autoencoder (SPA-CAE) model for feature extraction from Changshui Airport in Kunming QAR data. Provotar et al. [23] used LSTM layers in an autoencoder framework. Considering that, compared with normal data, abnormal data are difficult to be represented by low-dimensional feature vectors. By inputting the data into the LSTM autoencoder, the error of the AE reconstruction is used to judge whether the data are abnormal. These generative models hold great promise in the field of anomaly detection. However, these reconstruction-based models are difficult to capture long-range temporal dependencies and cannot explicitly address potential interactions between features. On the other hand, simply adding a network such as LSTM to a feedforward layer in AE or VAE does not perform detection well.

In summary, the information inside the window after dividing the window and the correlation between the window and the remaining time series are essential in anomaly detection. Although many approaches have been proposed, it is often impossible to achieve both. The correlation between windows is ignored and only one type of anomaly is detected. Based on these reasons, we propose a new VAE-LSTM hybrid deep model based on a multihead self-attention mechanism, which can effectively identify multiple types of anomalies without the limitation of window size.

3. Model

In this section, we introduce the overall workflow and internal structure of the VAE-based MHSA-LSTM hybrid model, as shown in Figure 2. We will introduce our model training process in an unsupervised way and explain the anomaly detection process on QAR data.

3.1. Problem Definition. A univariate time series is an ordered sequence of n real-valued variables arranged in chronological order. It can be formalized as $T = \{x_1, x_2, \ldots, x_n\}, T \in R$, where n is the length of the time series. Anomalies are observations or sequences of observations that deviate significantly from the general distribution of the data. In this paper, our goal is to discover outliers in QAR data through anomaly detection. Our method is divided into two parts: model training and anomaly detection. T as the training input can get a reconstructed sample T', calculate the anomaly score between T' and T, and compare it with the threshold to get the anomaly. Given a binary variable $y \in \{0, 1\}$, $y_t = 1$ is used to indicate that an anomaly occurred in the window of time t, and $y_t = 0$, no exception occurred.

3.2. Data Preprocessing. Data preprocessing is essential when building neural network models and can often determine the results of model training. First, we need to divide the given time series into a training set and a test set. A continuous data segment that does not contain anomalies is used as training data, and the rest with abnormal data is used as test data. Then, to improve the robustness of the model, we need to standardize the training set and test set. We first standardize the training set and then use the standardized parameters (mean and variance) of the training set to standardize the test set. The data standardization formula can be expressed as the following equation:

$$x' = \frac{(x - \mu)}{\sigma}, \tag{1}$$

where μ and σ are, respectively, represented as the mean and variance of the training set.

3.3. Training Model

3.3.1. Pretraining Using VAE Model. The VAE model is a typical generative model, which consists of two parts: an encoder and a decoder. First, we preprocess the input data X and send it to the encoder, which can encode higher data dimensions into a potential representation space Z, which is random and low-dimensional. The mean and variance of the output generate the corresponding latent variable z that satisfies the unit Gaussian distribution, so we can express the encoder as $q^{\phi}(z \mid X)$, and the parameter ϕ represents the mapping of the network from X to z. The other part of the

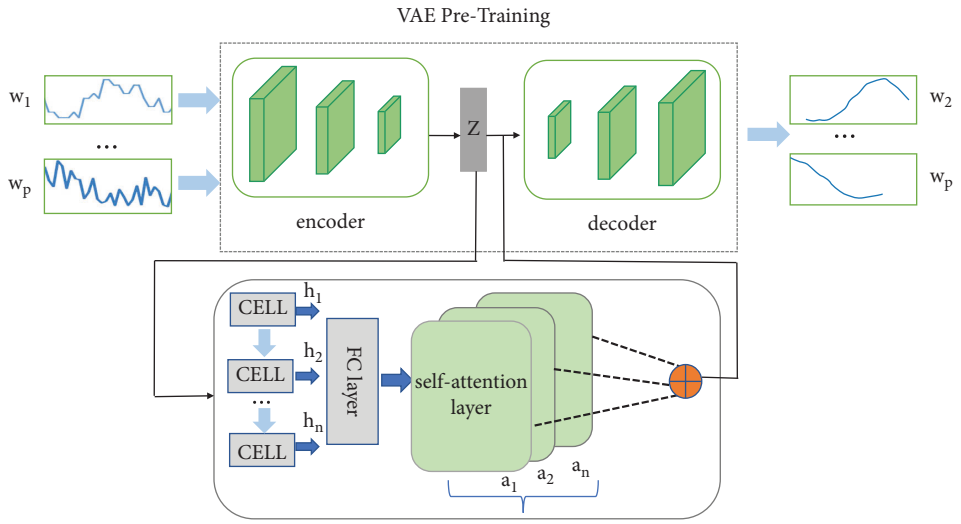


Figure 2: The workflow of the anomaly detection model.

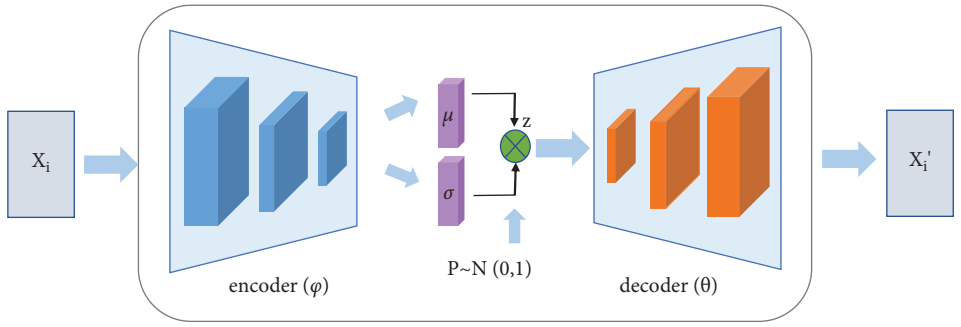


Figure 3: Architecture of variational autoencoder.

decoder of the VAE model can decode the latent variable z into the generated data x' which is similar to the real data and obeys the normal distribution with mean μ and variance σ, that is, $p \sim N(\mu, \sigma)$. Thus, we can express the decoder as $p^{\theta}(X \mid z)$, and the parameter θ represents the reconstruction of the network from z to X. Figure 3 shows the structure of the VAE network model.

In order to train the VAE model, we convert the training data into a local window as the input of the model, extract the features through the encoder, compress it into the latent space, and then reconstruct it. Given a time series $T = \{x_1, x_2 \ldots x_n\}$, where $x_i \in R$, each data point is the result of measurement at a characteristic time. To improve the accuracy of the model, we need to divide the entire time series into multiple subsequences, which are represented by time windows. We define w_t, a time window of length m at a given time t: $w_t = \{x_{t-m+1}, \ldots, x_{t-1}, x_t\}$. Because we use m data to predict the output, a total of $n - (m + 1)$ windows can be generated for training the VAE model. In this way, the time series T can be represented by the training input window sequence W: $W = \{W_1, W_2, \ldots, W_{n-m+1}\}$. After training, the model finally outputs the reconstructed window w_t' after the reconstruction of the window w_t through the decoder.

The loss function is the most basic and critical element used to measure the pros and cons of a model. The loss function of the VAE model is used to measure the information loss in the reconstruction process, and it is composed of the sum of the reconstruction error and the regularization term. Our VAE is trained with a loss function as shown in the following equation:

$$L(\theta, \phi) = -E_{z \sim q^{\phi}(z \mid x)}\left[\log p^{\theta}(x, z)\right] + KL\left(q^{\phi}(z \mid x) \| p(z)\right). \quad (2)$$

The first term is the reconstructed negative log-likelihood loss −ELBO (evidence lower bound), and the second term is the KL difference between $q^{\phi}(z \mid x)$ and $p(z)$. Our goal of training the VAE model is to minimize the sum of this reconstruction loss and KL divergence, which is equivalent to maximizing the ELBO loss to find the most suitable parameters θ and ϕ [24]. The objective function is the following equation:

$$\arg\max_{\theta, \phi} E_{z \sim q_{\phi(z \mid x)}}\left[\log p_{\theta(x \mid z)}\right] - KL\left(q_{\phi}(z \mid x) \| p_{\theta}(z)\right). \quad (3)$$

Through training, we optimized the parameters of the model while improving the loss, and the network finally converged, and a good generative model was obtained.

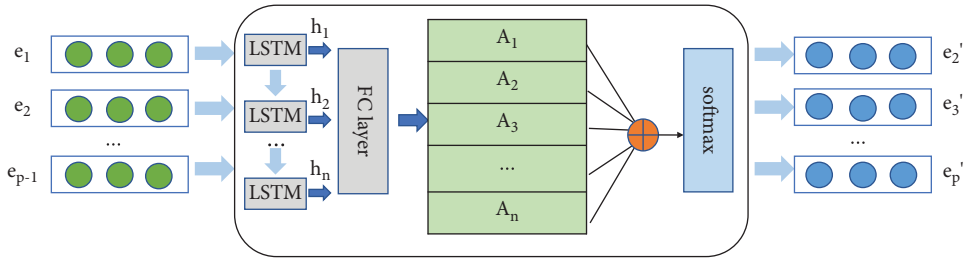


Figure 4: Architecture of the LSTM model with multihead self-attention.

3.3.2. LSTM Model Based on Multihead Self-Attention. A VAE model alone cannot achieve forecasting of time series because a VAE cannot encode or decode data outside the time window. Therefore, we use the LSTM model to act on the data after the dimension reduction of the VAE model, extract time features, and perform sequence prediction. We also introduced a multihead self-attention mechanism in the LSTM model to capture relevant information in different subspaces and highlight the importance of different features. The model structure diagram is shown in Figure 4. Below, we introduce the detailed information about the model.

(1) LSTM. The LSTM model is a type of recurrent neural network (RNN) that can solve the gradient descent or explosion problem that RNN may generate and learn and remember long-term relationships. Therefore, the LSTM network has achieved great success in time series data analysis [25]. The LSTM model is composed of LSTM memory cells. Each memory cell contains three gates with different functions, which are the input gate, output gate, and forget gate. These three gates are used to determine whether to accumulate or eliminate the information in the memory unit and to selectively retain the characteristics of the sequence. In this way, the network can determine the predicted output under this gating mechanism. Therefore, LSTM has become the basic framework for the task of processing sequential data with time information.

After pretraining the VAE model, we start to train the LSTM model. To prevent our model from overfitting, we divide the given training data into a sequence of p nonoverlapping rolling windows, which can be expressed as $W_t = [w_{t-(p-1)*m}, w_{t-(p-2)*m}, \ldots, w_t]$. Then, the window sequence W_t is encoded into a lower dimension by the encoder in the pretrained VAE model, and the output embedding can be expressed as $E_t = [e_1, e_2, \ldots e_p]$, where e_i represents the embedding of the i-th window in W_t. We train the encoder's output E_t as the input of the LSTM model and predict the next sequence e_2' based on the embedding e_1 of each window. Specifically, the LSTM model has n memory units, and each unit has a different set of internal weight parameters, namely, h and c. In each unit, there are two input data, respectively, the output and state $h_{(t-1)}$ and $c_{(t-1)}$ of the previous neuron and the input e_t of the current unit. Then, the hidden state of the output of the final unit can be expressed as the following equation:

$$h_t, c_t = \mathrm{LSTM}\left(e_t', h_{t-1}, c_{t-1}\right). \tag{4}$$

We express the hidden state of the embedded sequence e_t after passing through n LSTM units as the following equation:

$$H = \left(h_1, h_2, \ldots h_n\right). \tag{5}$$

(2) Multihead Self-Attention. Multihead self-attention is the core part of the transformer encoder-decoder model. It optimizes the traditional attention mechanism and greatly improves its performance. When performing feature extraction on a time series, you can focus your attention on a window sequence and assign weights to each time point of the sequence so as to determine the weight of their influence on the final output prediction results. An attention function is composed of a vector query, a key, and a value. The common attention mechanism is to make k and v equal to the input value, and q comes from the outside. After calculating the weight coefficients through the vectors q and k, the weighted summation with the vector v is performed to obtain the attention score. The self-attention mechanism obtains q, k, and v by making its own linear changes to the input value. Calculating the association between its own data is a feature extraction of the data itself. The calculation method of the self-attention mechanism is as shown in the following equation:

$$\mathrm{Attention}\left(Q, K, V\right) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{6}$$

where Q is the query, K is the key, V is the value, and d_k is the number of hidden units of the neural network. The multihead self-attention mechanism performs separate operations on the basis of the self-attention mechanism. Each head generates three vectors Q, K, and V through linear transformation and then performs self-attention calculations. Calculating once is a head, and calculating h times is the so-called long head. Finally, each head is spliced and converted into the same dimension as the input sequence. The formula is expressed as the following equations:

$$A_i = \mathrm{self} - \mathrm{att}\left(QW_i^Q, KW_i^K, VW_i^V\right), \tag{7}$$

$$\mathrm{Multihead}\left(Q, K, V\right) = \mathrm{Concat}\left(A_1, A_2, \ldots A_n\right)W. \tag{8}$$

We deploy it after the LSTM model, because when calculating each head, the parameters W after the linear transformation of Q, K, and V are different, which needs to be learned by the model. We use W_i to represent. The

attention layer takes the entire hidden state H as input and multiplies it with the parameter, W_i, to calculate the self-attention value of each head. The calculation formulas are shown as the following equations:

$$u_i = \tanh\left(W_i H + b_i\right), \tag{9}$$

$$v = \text{Multihead}\left(u_1, u_2, \ldots u_p\right). \tag{10}$$

After p operations, we join each operation result u_i to get a feature representation v. Finally, the obtained feature representation vector is sent to the softmax layer for prediction, and the prediction results are as shown in the following equation:

$$y = \text{softmax}\left(\text{wv} + b\right), \tag{11}$$

where w and b are the weight matrix and bias of the final linear layer. Finally, we train our model by minimizing the error between the original data and the predicted data.

3.4. Anomaly Detection. Our anomaly detection method is divided into three stages: preprocessing, training, and detection. Among them, the training and detection stages share the first data preprocessing stage, and the data are standardized and divided into time windows of length m. After training, our model can be used for anomaly detection. First, we input the preprocessed test set sequence W_t into the LSTM model, which represents the pm data contained in time t. Then, we use the pretrained VAE model to reduce the dimensionality of W_t and encode W_t into a low-dimensional space by extracting features to obtain an embedding sequence E_t. The coded representation is used in the prediction stage of the LSTM model. The LSTM model predicts the next embedding e_i by learning $e_{(i-1)}$, as shown in the following equation:

$$e_i' = \text{LSTM}\left(e_{(i-1)}\right). \tag{12}$$

Finally, we use the decoder of the VAE model to perform feature restoration and reconstruct the predicted e_i' into a new window $W_{t-(p-i)*m}$, which is as shown in the following equation:

$$w_{t-(p-i)*m'} = \text{Decoder}\left(e_i'\right), \tag{13}$$

where $i \in \{2, 3, \ldots p\}$.

In the anomaly detection stage, our model will get a total of two results, which are the predicted value calculated based on the prediction model and the reconstructed value obtained based on the reconstruction model. We measure the degree of anomaly by calculating the root mean square error (RMSE) between the reconstructed window and the original window as the anomaly score of the window. The higher the abnormality score, the greater the possibility of abnormality. The formula to calculate the RMSE error is as shown in the following equation:

$$\text{score} = \sqrt{\sum_{i=2}^{p}\left(w_{t-(p-i)*m'} - w_{t-(p-i)*m}\right)^2 / p}. \tag{14}$$

Among them, $w_{t-(p-i)*m}$ is the true value and $w_{t-(p-i)*m'}$ is the reconstruction value. The calculated result is the sum of the reconstruction errors of each time step of the entire window. The sum of the errors of these data points can be used as the anomaly score of the entire window. In order to effectively detect anomalies, we also need to set a threshold θ on the anomaly score. If the anomaly score is higher than this threshold, we will regard the window sequence as a window where anomalies may occur.

For this kind of binary classification problem, it is essential to choose an appropriate threshold, which can maximize the performance of the classifier. Some commonly used methods of threshold selection include artificially setting a fixed threshold. When the reconstructed value is greater than (or less than) the fixed threshold, it is judged that the value is abnormal. There are also some models that detect anomalies through the 3-sigma method. Standard deviation is a commonly used quantitative form that reflects the degree of data dispersion, and the dispersion is the most basic and important indicator for evaluating the quality of a method. Therefore, when the outlier exceeds 3 times the standard deviation, it can be regarded as an outlier. The advantage of these methods is simplicity, but obviously, the solution of setting a fixed threshold during deployment is not enough, and it is prone to false positives and underreports, and the scene adaptability is low. In order to avoid the above situations and better illustrate our model, we use a method of maximizing the $F1$ metric to automatically select the best threshold. The $F1$-score value is the harmonic average of the precision rate and the recall rate, and the accuracy and recall rate of the model can be considered at the same time in the detection. It can be calculated by the following equation:

$$F1 = 2 * \frac{(P * R)}{(P + R)}. \tag{15}$$

In the formula, P represents the accuracy rate of the detection model, and R represents the recall rate of the detection. First, we compute the reconstruction error RMSE for each window as the anomaly score and then compute the $F1$-score for multiple thresholds using an iterative grid search between the minimum and maximum reconstruction errors. We record the selected threshold θ when the $F1$-score value is the highest and use it as the optimal threshold. Any sequence of windows above this threshold will be considered anomalous. Because the number of anomalies in QAR data is low, we mainly focus on continuous anomalies or anomalous segments. If any point in the anomaly segment is correctly detected, all points in the anomaly window are identified as true positives, and the others are considered normal.

4. Experimental Evaluations

In this section, we conduct several comparative experiments to demonstrate the effectiveness of our method from different perspectives. We first introduce the real-world QAR dataset used in the experiments, evaluate the performance of our model on the dataset, and compare it with

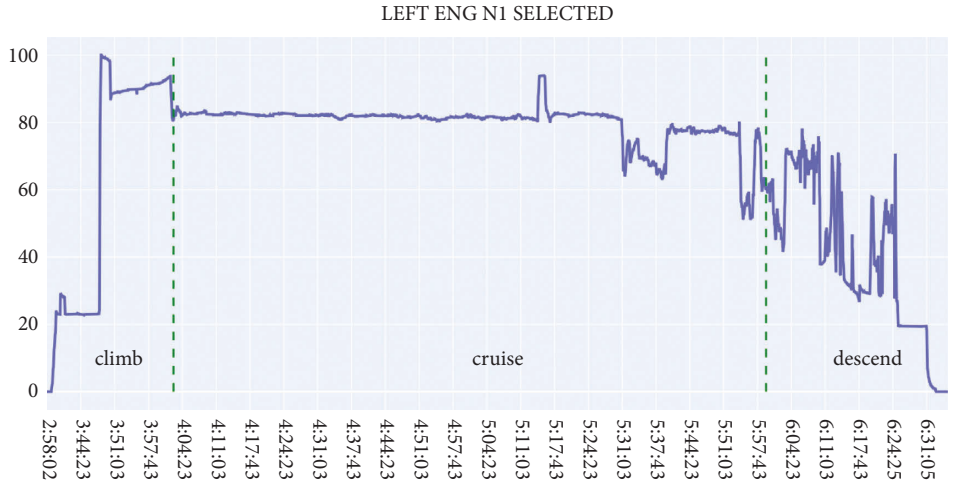


Figure 5: The $N1$ engine parameter of QAR.

Table 1: Comparison of anomaly detection performance based on precision, recall, and F1-score.

Methods	Climb			Cruise			Descent		
	P	R	$F1$	P	R	$F1$	P	R	$F1$
IF	0.5833	0.3293	0.4211	0.6333	0.7039	0.6667	0.5108	0.7055	0.5926
LSTMS	0.6944	0.9801	0.8195	0.4425	0.9440	0.6136	0.8240	0.6352	0.7769
LSTM-AE	0.8885	0.9426	0.9147	0.8768	0.9417	0.9134	0.7284	0.8534	0.7860
LSTM-VAE	0.7722	0.9443	0.8496	0.8119	0.9716	0.8961	0.8902	**1.0**	0.9419
VAE-based MHSA-LSTM	0.9145	0.9833	0.9503	0.8840	**1.0**	0.9384	0.9453	**1.0**	0.9718

other state-of-the-art methods (4.1). Second, we analyze how different parameter sizes affect the performance of the method (4.2). Then, we analyze the time performance of the model (4.3), and finally, we evaluate our algorithm (4.4) on different real-time series datasets.

4.1. The Comparisons with Different Methods

4.1.1. Datasets. Each QAR data file records the whole process of an aircraft from take-off to landing, including 11 stages, TAXI OUT, TAKE OFF, 2 SEGMENT, INI. CLIMB, CLIMB, CRUISE, DESCENT, APPROACH, FINAL, LANDING, and TAXI IN. These stages can be generally divided into three processes: the climb, cruise, and descent of the aircraft. Figure 5 shows the thrust parameters of the $N1$ engine generated by an aircraft of an airline during a voyage. It can be clearly seen that the aircraft tends to climb, stabilize, and then descend during the voyage.

Since the amplitude and speed of the data changes in different flight stages of each flight aircraft are different, we adopt a segmentation method. We divide the entire data into three segments: climate, cruise, and descend, according to the flight stage parameter FLIGHT_PHASE, and pass through each stage, respectively. The sliding window extracts local features for anomaly detection. Segmentation not only reduces the dimension of the data but also reduces the amount of computation and enhances the adaptability of the algorithm to QAR data. In this experiment, the $N1$ parameters generated by 100 normal flights of the same aircraft in the real world are selected, and each segment is connected to a file for training and anomaly detection after segmentation, and the data will have obvious circularity. There are a total of 82,606 sampling values in the climbing stage; 107,758 sampling values in the cruise stage; and 69,330 sampling values in the descending stage.

4.1.2. Experimental Setup. Our proposed method is mainly implemented by the Python programming language. It uses the well-known Tensorflow and Keras deep learning frameworks and includes multiple statistics and visualization packages, including Sckit-learn, Pandas, and Numpy. For the hyperparameters used in the model training process, we set the hidden size of the LSTM unit to 64 by default; $h_{\dim}$ for dimension of the hidden layer in the VAE model is set to 512, and $z_{\dim}$ for dimension of the latent variable Z is set to 10; the number of heads n of multihead self-attention is set to 6; and the number of samples for each training batch size is set to 64. In the training details, the learning rate of the VAE model and LSTM model is set to 0.0002; adaptive moment estimation (Adam) is used as the optimizer to optimize the gradient. The reconstruction loss of the mean square error MSE serves as the loss function of the LSTM model, while the loss function of the VAE consists of the reconstruction loss of the mean square error MSE and the Kullback-Leibler divergence loss of the difference between the target distributions. The model is trained for 50 epochs. For the other comparison models, we also use the hyperparameters described above. All models that require sliding windows are compared under the condition that the default window length is 144. When testing each model, we retained the results with the highest $F1$-score.

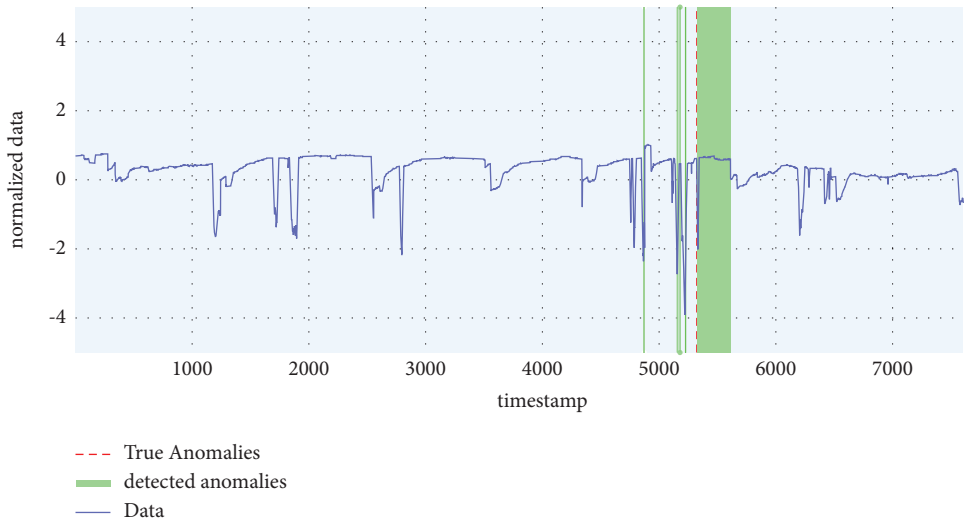


Figure 6: Anomaly detection in the climb phase dataset.

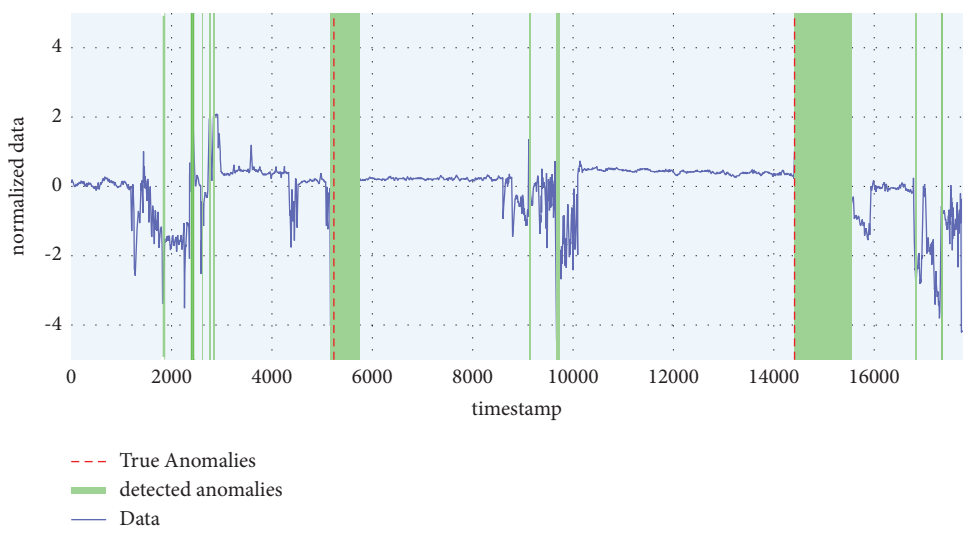


Figure 7: Anomaly detection in the cruise phase dataset.

4.1.3. Evaluation Metrics. The performance indicators used in the comparison experiments are precision, recall, and $F1$-score, which are commonly used evaluation indicators in anomaly detection. Equation (15) already gives the definition of the $F1$-score. Equations (16) and (17) give the definitions of precision and recall.

$$\text{precision} = \frac{TP}{(TP + FP)}, \tag{16}$$

$$\text{recall} = \frac{\text{TP}}{(TP + FN)}, \tag{17}$$

where TP stands for true positives, FP stands for false positives, and FN stands for false negatives. When a window is detected and marked as abnormal, where TP is the number of correctly detected abnormal points, FP is the number of normal points that are incorrectly predicted as abnormal points, and FN is the number of abnormal points that are incorrectly predicted to be normal. Accuracy is the ratio of the number of correctly predicted samples to all predicted samples of a particular class and can be used to measure the quality of model prediction. Recall is calculated as the ratio of correctly predicted samples to the total number of instances of the same type. The higher the recall, the easier it is for the model to detect anomalies. Higher recall is very important. Precision is generally paired with recall to evaluate model performance, but sometimes there are contradictions. Therefore, in order to have a more comprehensive evaluation of anomaly detection, we more comprehensively consider the $F1$-score, which is the harmonic mean of precision and recall.

4.1.4. Results. To demonstrate the overall performance of our proposed method, we compared it with four other unsupervised anomaly detection models. They are isolation forest (IF) [26], long-short-term memory (LSTMS) [16],

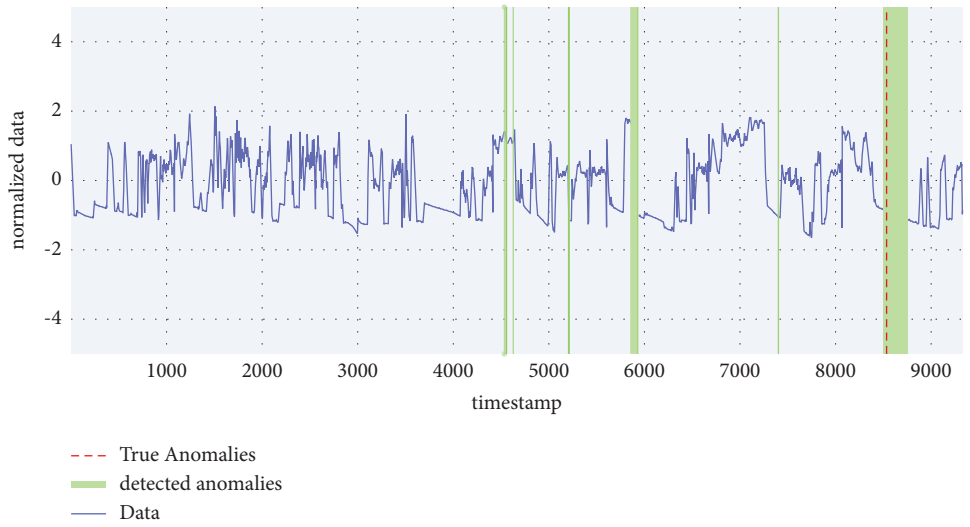


Figure 8: Anomaly detection in the descent phase dataset.

Table 2: The effect of head number.

Heads	Climb			Cruise			Descent		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
2	0.8119	0.9964	0.8947	0.8584	0.9826	0.9238	0.8511	1.0	0.9196
4	0.8299	0.9960	0.9054	0.8957	0.9803	0.9449	0.9127	0.9987	0.9543
6	0.9145	0.9890	0.9503	0.8782	1.0	0.9356	0.9265	1.0	0.9618
8	0.8465	1.0	0.9119	0.8934	0.9763	0.9437	0.8707	0.9981	0.9309

Table 3: Training time per epoch (min).

Methods	Climb	Cruise	Descent
LSTM-AE	5.05	7.57	4.27
LSTM-VAE	2.04	1.69	2.58
VAE-based MHSA-LSTM	1.75	1.32	1.2

LSTM-AE [27], and LSTM-VAE [28]. We report for each method the results associated with the highest *F*1-score values. Table 1 details the performance results of all methods on the climb, cruise, and descent datasets. The results show that our model significantly outperforms other methods in precision, recall, and *F*1-score on all datasets, where the precision is able to improve by 0.5–0.7. It can also be observed that our model performance is well-balanced across different stages of QAR data. Figures 6–8 show the visualization results of our method for anomaly detection on the climb, cruise, and descent datasets. From these figures, we can see that our method can correctly find the time window in which abnormal events occur, which proves that our model has a high recall rate. The very few false positives plotted in the graph are because, historically, such a spike has been infrequent, so it was detected as an anomaly by our model. We may need more domain knowledge to solve this problem in the future.

The methods we compare include the machine learning method IF, the traditional predictive model LSTM, and the combination of LSTM with autoencoders and generative models. It can be observed that the IF method performs the worst. IF builds a collection of iTrees for a given dataset, and then the instances go through all iTrees. The anomaly score is the average of all path lengths. It does not observe time information. In time series, time correlation is essential. The prediction model composed of LSTMs may lead to large deviations in the results due to the uncertainty of the prediction results. Thus, the results of precision and recall are relatively low. The autoencoder reconstructs time series through an encoder-decoder framework. On this basis, LSTM is combined with the autoencoder, and the encoder and decoder of AE are composed of multiple LSTM units. The main role of AE is to reduce the dimensionality of the data, form a low-dimensional latent vector, and combine it with LSTM to capture the long-term correlation of time series. However, in contrast, as a generative model, VAE can generate new data completely different from the training data through training and satisfy the standard normal distribution. It can be seen from the experimental results that the combination of VAE and LSTM is much better than AE. Detection performance improved, but significant performance fluctuations were seen between different stage datasets. Our method adds a multihead self-attention mechanism on top of this and calculates the dependencies between long-distance windows separately through multiple heads. The weighting calculation is applied to the reconstruction of the VAE decoder. Therefore, our model captures the long-term dependencies of time series more easily than other methods. The results also show that our model achieves 100% recall on the cruise and descent datasets. This

Table 4: Statistical information of four public benchmark datasets.

Dataset	Total length	Train length	Test length	Mean	Std.	Anomaly rate (%)
KPI1	90000	75000	15000	2.3840	0.9174	5.18
KPI2	17562	10000	7562	0.1911	0.1004	0.79
NAB1	18050	15500	2550	37.4794	14.4096	0.08
NAB2	4032	3000	1032	45.1079	1.8774	0.29

Table 5: Anomaly detection performance on four public benchmark datasets.

Methods	KPI1			KPI2			NAB1			NAB2		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
LSTMS	0.7639	0.6544	0.7049	0.5850	0.9997	0.7382	0.4536	**1.0**	0.6241	0.8604	0.7915	0.8779
LSTM-AE	0.7261	0.8521	0.7841	0.6773	0.8230	0.7430	0.7611	0.6807	0.6870	0.7627	**0.8733**	0.8142
LSTM-VAE	0.7815	0.9545	0.8594	0.8734	0.9271	0.8995	0.7468	**1.0**	0.8550	0.9090	0.6563	0.7623
VAE-based MHSA-LSTM	**0.8221**	**1.0**	**0.9023**	**0.8786**	**1.0**	**0.9354**	**0.8731**	**1.0**	**0.9322**	**0.9547**	0.8146	**0.8791**

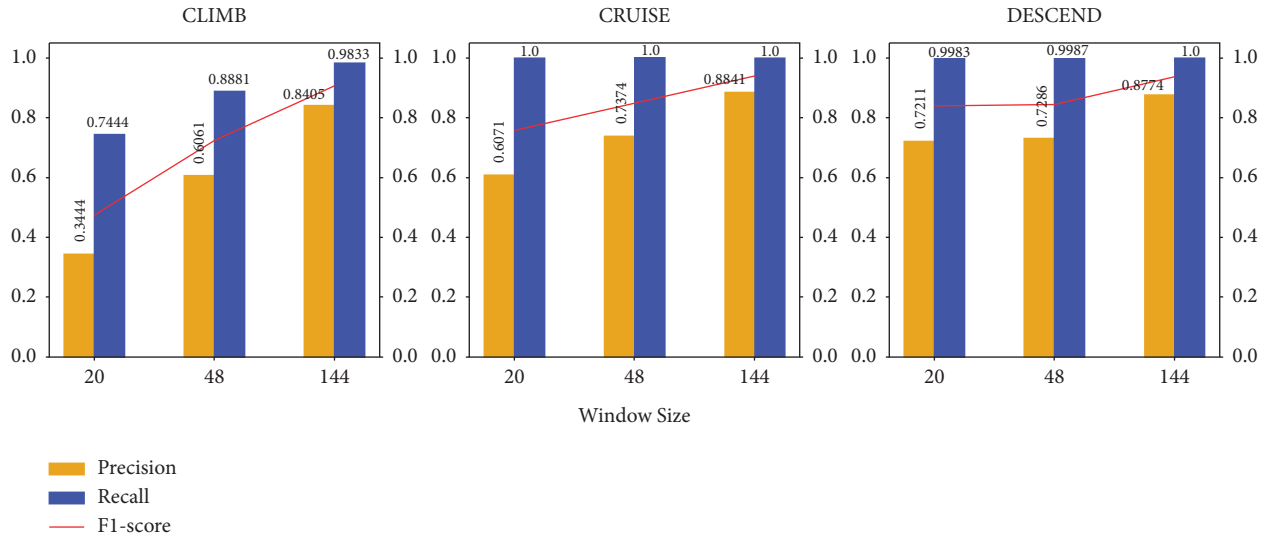


Figure 9: Comparisons with different window sizes.

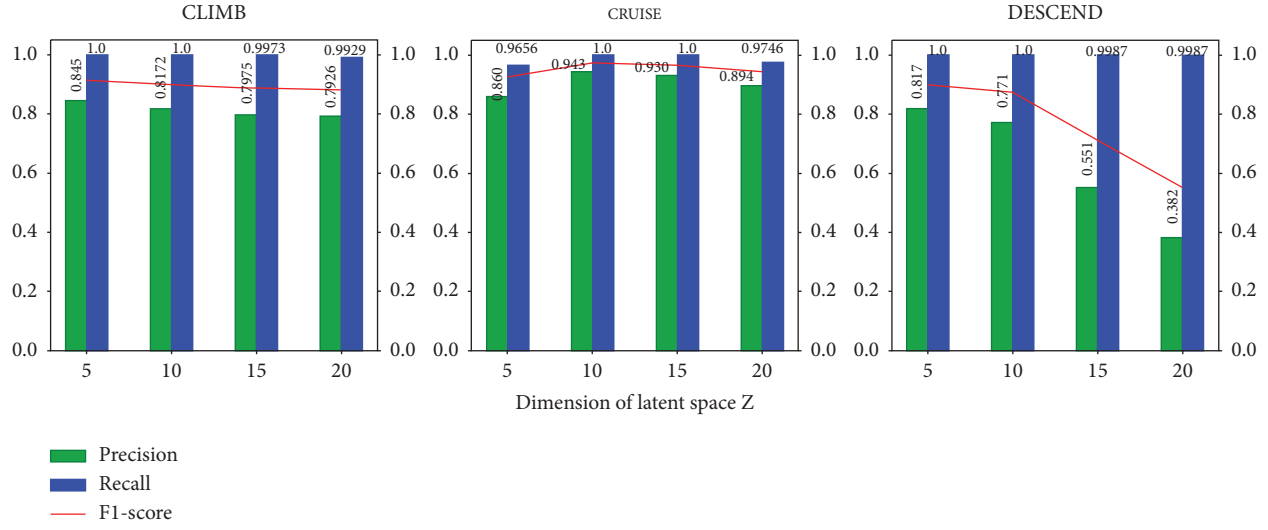


Figure 10: Comparisons with different z_dim.

shows that we have no abnormal points that are wrongly predicted to be normal and can effectively detect both short-term and long-term anomalies. Overall, our method shows better performance than other methods.

4.2. Effect of Parameters. In this section, we investigate the different effects of different parameters and factors on the method's performance, and all experiments are done using the three datasets of QAR.

4.2.1. Effect of Different Window Sizes. The first factor is the different window sizes in different datasets. The window size has an impact on the results of anomaly detection, because it not only affects the speed and efficiency of anomaly detection but also directly affects the detection accuracy. It is crucial to model the data within the window interval by choosing the appropriate window size for different datasets. We set the window size to 20, 48, and 144 for the experiments, and other parameters remained the same. The results are shown in Figure 9. From the results, we can observe that on all datasets, when the window size is increased, higher precision, recall, and $F1$-score can be obtained. This means that if the duration of the window is too short, the model may fail to learn that long-term anomalies have occurred. In QAR data, anomalies that occur during flight are more likely to be continuous segments than isolated points. This proves that our model structure can detect abnormal events for a longer period of time, and the data are relatively stable in the climb and cruise stages, which makes it more suitable for relatively large size windows to improve the detection efficiency.

4.2.2. Effect of Latent Variable z Dimension. In addition to the window length, we also investigate the link between $z_{\dim}$ for the dimension of latent variable z and detection performance. In VAE, the dimension of the latent variable space is a crucial parameter, which represents the important information required in the original data and can determine the representation ability of the latent space. VAE uses a probability distribution over the latent space to sample new data that can represent the characteristics of the original data. The embedding results obtained by sampling in different dimensions are different, and the reconstructed data are also very different. We set the dimensions of the latent variable z to 5, 10, 15, and 20 to observe its performance impact on the anomaly detection reconstruction process. Figure 10 shows the experimental results. The results show that if the latent variable z is located in a very large dimension, it will cause unnecessary redundancy to hinder the learning of the model, which may lead to the performance degradation of the VAE model training data. However, this does not mean that the smaller the latent variable space, the better. Considering that there is a special case, when the dimension is too small, VAE will lose a lot of information in the encoding stage and cannot decode. The model cannot fully capture the time dependency, resulting in poor model performance. It can be seen from the figure that the $F1$-score is relatively stable when the dimension is moderate. This confirms the above discussion. A suitable latent space size can make the model more robust in anomaly detection.

4.2.3. Effect of Head Number in MHSA Mechanism. In order to explore the effect of the number of heads on the model performance in the multihead self-attention mechanism, we set different head numbers of 2, 4, 6, and 8 for experiments. The experimental results are shown in Table 2. The results showed that in the climb and descent stages, the $F1$-score was the highest when head = 6. In the cruise phase, the $F1$-score is the highest when head = 4. Overall, the performance of the model fluctuates. As the number of heads increases, each head captures different aspects of information, and the model can capture more temporal information. The model performs the worst when there are only 2 heads, but an excessive number of heads makes the information captured between each self-attention head redundant, which weakens the model's ability to extract effective correlations. Combining the experimental results and efficiency, we set the number of heads to 6 in our implementation.

4.3. Analysis of Training Time. In this subsection, we also record the running time of epochs in each stage dataset and compare our method with several other deep learning hybrid models. All methods are compared on the same system. Table 3 shows the results obtained. The results show that our model is less time-consuming than other models, because we added a multihead self-attention mechanism to the LSTM. The parallel operation of multiple self-attention mechanisms can not only extract hidden features at a deeper level but also reduce the dimension and the amount of calculation. Therefore, we not only achieved good performance in anomaly detection but also reduced training time and improved operating efficiency.

4.4. The Comparisons of Using Different Datasets. In this subsection, to verify the feasibility of our method, we conduct experiments on several different public benchmark datasets. They are the KPI and NAB datasets that are often used to perform experiments in time series anomaly detection. Normal and abnormal are already marked in these datasets. The KPI dataset is from the AIOps Challenge held by Tsinghua University in 2018 [29]. Many Internet companies monitor the data generated by various performance indicators in order to ensure the stability of web services, such as CPU usage and server health, and other performance indicators. We randomly selected two time series from the KPI dataset for experiments. The NAB dataset, provided by artificial neural network company Numenta, contains a variety of streaming data in real-time applications, consisting of multiple labeled real-world and artificial time series data files. We selected the CPU usage of Amazon Web Services (AWS) servers and AWS EC2 servers collected by the Amazon Cloudwatch service as our dataset. Table 4 lists the data such as size, mean, standard deviation, and anomaly ratio of the four datasets, and it can be seen that these four datasets are significantly different.

We divided each data set into two parts: training set and test set, because our model needs to use normal data to train, so we removed the abnormality in the training data and got normal data. Outliers in the test set are reserved for testing.

Table 5 shows the experimental results. It can be clearly seen that our method outperforms other methods on these four public datasets. The accuracy of our model on these datasets is different. The $F1$-score of most datasets is above 0.9, and most datasets have achieved a 100% recall rate, which indicates that the number of false negatives (FN) is low. Because of the diversity of KPI and NAB datasets, some are cyclical, and some are unstable and fluctuating. This proves that our method performs well, can also detect different types of data anomalies, and has good generalization ability.

5. Conclusion

In this paper, we propose VAE-based MHSA-LSTM, an unsupervised deep learning-based method for anomaly detection in time series. The method can be divided into two stages. One is the model training stage. First, the variational autoencoder model is pretrained, and the features of normal data are learned, which can form stable local features in each window. The second is the anomaly detection stage, which uses the learning ability of the LSTM model for temporal representation and the feature extraction ability of the self-attention mechanism to identify anomalies based on the anomaly scores of the sample reconstruction calculation window. The VAE-based MHSA-LSTM combines encoder-decoder, generator, and multihead self-attention mechanism, which can detect all types of anomalies more comprehensively, quickly, and accurately. In the experimental part, we apply VAE-based MHSA-LSTM to the QAR dataset generated by real-world flights. Compared with several other classical reconstruction-based time series anomaly detection methods, the results show that our method has a better effect. In addition, we also applied our method on other public datasets with stable results.

Although our method achieves good performance and can accurately detect anomalies, there are still some limitations. Our model needs to be trained on the training data before anomaly detection, and the training set must ensure that there is no abnormal data. This presents some difficulties with the collection and processing of data. Therefore, in the future, we will explore the space for further development based on some of the ideas presented in this article.

Data Availability

The data used in this study are available from the corresponding author upon request.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Acknowledgments

This work was supported by the project of Natural Science Foundation of China (Nos. 61402329 and 61972456) and the Natural Science Foundation of Tianjin (Nos. 19JCYBJC15400 and 21YDTPJC00440).

References

[1] K. Mitchell, B. Sholy, and J. Alan, "General aviation aircraft flight operations quality assurance: overcoming the obstacles," *IEEE Aerospace and Electronic Systems Magazine*, vol. 22, no. 6, pp. 9–15, 2007.

[2] H. Heng, J. Zhang, and C. Xin, "Research on aircraft engine fault detection based on support vector machines," in *Proceedings of the International Conference on Consumer Electronics, Communications and Networks*, pp. 496–499, CECNet, Yichang, China, April 2012.

[3] T. Ergen and S. S. Kozat, "Unsupervised anomaly detection with LSTM neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 3127–3141, 2020.

[4] Bo Zong, S. Qi, R. Martin et al., "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proceedings of the International Conference on Learning Representations*, Vancouver, BC, Canada, May 2018.

[5] D. Park, Y. Hoshi, and C. C. Kemp, "A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1544–1551, 2018.

[6] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the Annual Conference on Neural Information Processing Systems*, pp. 5998–6008, Long Beach, California, USA, December 2017.

[7] M. Braei and S. Wagner, "Anomaly detection in univariate time-series: a survey on the state-of-the-art," *CoRR*, 2020, https://arxiv.org/abs/2004.00433.

[8] Q. Yu, L. Jibin, and L. Jiang, "An improved arima-based traffic anomaly detection algorithm for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, 2016.

[9] T. B. Ottosen and P. Kumar, "Outlier detection and gap filling methodologies for low-cost air quality measurements," *Environmental Science-Processes & Impacts*, vol. 21, pp. 701–713, 2019.

[10] G. Münz, Sa Li, and G. Carle, "Traffic anomaly detection using k-means clustering," *GI/ITG Workshop MMBnet*, vol. 7, p. 9, 2007.

[11] L. Li, S. Das, R. John Hansman, R. Palacios, and A. N. Srivastava, "Analysis of flight data using clustering techniques for detecting abnormal operations," *Journal of Aerospace Information Systems*, vol. 12, no. 9, pp. 587–598, 2015.

[12] W. Zhao, L. Li, S. Alam, and Y Wang, "An incremental clustering method for anomaly detection in flight data," *CoRR*, vol. 132, Article ID 09874, 2020.

[13] C. Zeng, R. Wang, and Q. Zuo, "Analysis of abnormal flight and controllers data based on dbscan method," *Security and Communication Networks*, vol. 2022, Article ID 7474270, pp. 1–8, 2022.

[14] E. Smart, D. J. Brown, and J. Denman, "A two-phase method of detecting abnormalities in aircraft flight data and ranking their impact on individual flights," *IEEE Transactions on*

Intelligent Transportation Systems, vol. 13, no. 3, pp. 1253–1265, 2012.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in *Proceedings of the International Conference on Knowledge Discovery & Data Mining*, pp. 387–395, London, UK, July 2018.

[17] A. Khorram, M. Khalooei, and M. Rezghi, "End-to-end cnn+lstm deep learning approach for bearing fault diagnosis," *Applied Intelligence*, vol. 51, no. 2, pp. 736–751, 2021.

[18] A. Vahdat and J. Kautz, "NVAE: a deep hierarchical variational autoencoder," *Annual Conference on Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 19667–19679, 2020.

[19] D. Li, D. Chen, B. Jin, S. Lei, G. Jonathan, and N. G. See-Kiong, "MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks," in *Proceedings of the International Conference on Artificial Neural Networks*, pp. 703–716, Lausanne Switzerland, September 2019.

[20] K. Kishore, S. Sarkar, V. Venugopalan, and M. Giering, "Anomaly detection and fault disambiguation in large flight data: a multi-modal deep auto-encoder approach," *Annual Conference of the PHM Society*, vol. 8, 2016.

[21] M. Memarzadeh, B. Matthews, and I. Avrekh, "Unsupervised anomaly detection in flight data using convolutional variational auto-encoder," *Aerospace*, vol. 7, no. 8, 2020.

[22] Q. Wang, K. Qin, B. Lu, and R. Huang, "Feature extraction of qar data via sequence-parameter attention based convolutional autoencoder model," in *Proceedings of the IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, pp. 352–355, IEEE, Changsha, China, December 2021.

[23] O. I. Provotar, Y. M. Linder, and M. Maksym, "Unsupervised anomaly detection in time series using lstm-based autoencoders," in *Proceedings of the IEEE International Conference on Advanced Trends in Information Theory*, pp. 513–517, Kyiv, Ukraine, December 2019.

[24] D. Jimenez Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proceedings of the International Conference on Machine Learning*, pp. 1278–1286, PMLR, Beijing, China, June 2014.

[25] Y. Wu, M. Schuster, Z. Chen et al., "Google's neural machine translation system: bridging the gap between human and machine translation," *CoRR*, 2016, https://arxiv.org/abs/1609.08144.

[26] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008)*, pp. 413–422, IEEE Computer Society, Washington, DC, USA, December 2008.

[27] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, "Lstm-autoencoder based anomaly detection for indoor air quality time series data," *CoRR*, 2022, https://arxiv.org/abs/2204.06701.

[28] S. Lin, R. Clark, and R. Birke, "Anomaly detection for time series using vae-lstm hybrid model," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4322–4326, IEEE, Barcelona, Spain, https://ieeexplore.ieee.org/author/37297514400, Barcelona, Spain, May 2020.

[29] N. Zhao, J. Zhu, Y. Wang et al., "Automatic and generic periodicity adaptation for kpi anomaly detection," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 1170–1183, 2019.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 5343909, 9 pages
https://doi.org/10.1155/2022/5343909

Research Article

Medium and Long-Term Fault Prediction of Avionics Based on Echo State Network

Chi Gao, Bin Li, and Zhen Dai

AVIC Chengdu Aircraft Design & Research Institute, Chengu 610000, China

Correspondence should be addressed to Chi Gao; 160908240@stu.cuz.edu.cn

Received 8 July 2022; Revised 31 August 2022; Accepted 5 September 2022; Published 16 September 2022

Academic Editor: Chin-Ling Chen

As one of the core equipment of aircraft, avionics provide a power source for flight. Avionics are complex and highly susceptible to environmental factors. Failure in the long flight process is also relatively large, affecting the stability and safety of aircraft operation. Therefore, it is of great significance to predict the typical faults of avionics. At present, a lot of research achievements have been made on the fault diagnosis of avionics, but the failure prediction of avionics is rarely involved, especially the middle and long-term fault prediction. Hence, this paper proposes a fault prediction method for avionics based on an echo state network. In particular, one-dimensional wavelet denoising filtering and z-score standardized preprocessing are carried out to obtain pure useable data first. Then, the set training data are input into the ESN model. When the model is well trained, the test data can be used to test the model. Finally, the experimental results demonstrate that the proposed ESN model can effectively improve the medium and long-term prediction accuracies of the faults in avionics equipment. Besides, the proposed model can not only identify the types of faults but also predict the specific time when the faults occur. It guarantees the safe and stable operation of the equipment and supports the stable development of the air transport industry, which has great theoretical and practical application value.

1. Introduction

The rapid development of China's economy has promoted the leaps and bounds development of China's aviation field. Avionics provide continuous power output for civil and military aircraft. They are the core part of the aircraft and should be used to the maximum extent possible to ensure their normal operation [1, 2]. If a failure occurs, it is bound to cause economic and personnel safety losses, but the occurrence of failure always exists; analysis and research for the typical failure of avionics equipment can predict the failure or anomaly in advance and, to a certain extent, can reduce the loss to the minimum. The failure of avionics not only causes huge economic losses but also may affect national security. So, it is very important to analyze the reliability of avionics. At present, electronic equipment is more and more widely used in aerospace products, facing more and more severe working environments. An important goal of electronic equipment structure design is how to ensure high reliability in order to ensure that electronic equipment is in a vibration and impact environment with high reliability [3]. Modern military aircraft need to perform a variety of complex tasks, and the onboard equipment needs to ensure that it can work in different complex environments, which makes the safety and reliability of the electronic equipment on the aircraft become more and more demanding. As a result, airborne equipment is prone to failure, especially airborne electronic equipment. For example, when the electronic equipment is overworked, the aircraft may break down during the flight, thus affecting the overall performance of the aircraft or even causing serious consequences. However, in actual flight, avionics faced a more complex environment, such as the excitation of its own engine and the disturbance of external pressure, which have a great influence on its reliability [4, 5].

The manufacturing industry and the electronic equipment in the manufacturing industry are also developing in the direction of more intelligence, so the problem of the time prediction of the failure of electronic equipment arises. From the technical difficulty, due to the rapid development of 'China's manufacturing industry and the automated electronic equipment industry, the demand for the manufacturing industry is growing [6, 7]. In order to meet the demand of the manufacturing industry and other industries in automatic electronic equipment, the internal structure of automatic electronic equipment is gradually becoming more and more complicated, advanced, and intelligent. Suppose there is no special technical personnel to maintain the electronic equipment once the electronic equipment fails. First, it will lead to a longer maintenance cycle of electronic equipment, affecting the normal use of electronic equipment. Second, because of the unnecessary inspection, repeated inspection, and incorrect inspection of electronic equipment, it is easy to cause the under-maintenance and over-maintenance of electronic equipment, which leads to an increase in maintenance costs and waste. From the perspective of cost, on the one hand, as the structure of electronic equipment becomes more and more precise and its functions become more and more intelligent, the price of purchasing electronic equipment will become more expensive. On the other hand, when professional and technical personnel are hired to maintain electronic equipment, the cost will become higher. From the point of view of maintenance, maintenance can be divided into maintenance after failure and periodic maintenance of electronic equipment. For the maintenance after the failure, when the maintenance is not timely, the possible cost loss cannot be measured. When the maintenance is timely, it will also cause the breakdown of the whole electronic equipment due to the removal of parts. For periodic maintenance, although it can avoid the adverse impact of failure, it will carry out a lot of unnecessary periodic maintenance of electronic equipment [8, 9].

In addition to ensuring basic safety issues, the health status of avionics systems needs to be provided in real time, so that faults can be predicted and located accurately and quickly in a short time once they occur. In addition, if the failure can be predicted successfully, it can provide a reference and basis for avionics maintenance and improve the safety and reliability of equipment operation. Therefore, it is extremely urgent to improve the safety and reliability of avionics and reduce the occurrence of major accidents. This is not only an important way for various small and medium-sized enterprises to improve their competitiveness but also an inevitable trend of the entire aerospace system. And because of the increasing intelligence of modern weapons and equipment, the Prognostics and Health Management (PHM) system has been developed in the United States and other countries. It uses sensor information and expert diagnosis, intelligent inference model, and fault prediction algorithm. The maintenance capability and fault prediction capability of aviation system is one of the main research directions of PHM [10, 11].

Through reading relevant literature, we know that in recent years, the prediction of failure time of electronic equipment has become a hot research topic. Often, people do not know when the next device failure will occur, and it takes more labor and resources to find the fault. Therefore, through reasonable technical means to predict and accurately maintain the failure of aerospace electronic equipment, not only the service life of electronic equipment can be prolonged but also the safety requirements of aviation electronic equipment can be met. How to deal with time series is very important in fault prediction technology, such as predicting the time when the fault occurs. Time series data is the trend of some variables changing with time, and processing time series is to find out these trends for predictive analysis [12, 13]. Time series prediction means that for those complex system whose precise mechanism model cannot be established, it takes the experimental or observed multivariate time series as the entry point to study the internal change rules of the system and predict the future changes of the system. Before analyzing the time series, the data are preprocessed, including smoothing, denoising, and removing outliers. In the global prediction model, all observation samples are regarded as research objects, and the dynamic persistence of the unknown system is studied by establishing the corresponding nonlinear mapping relationship.

Traditional global prediction models mainly use quotient polynomials to achieve global approximation. Local prediction models can be divided into linear and nonlinear models according to different model properties. Compared with the global prediction model, the local prediction model has the advantages of a simple mathematical model, fast training speed, no complicated parameter estimation, and wide application. The above global prediction model and local prediction model are established under the condition of relatively complete observation data. If the avionics operation data are missing or have time-varying characteristics, the prediction results will be greatly affected. The adaptive fault prediction model is a kind of separation and demand method which appears in recent years. Because this method can adjust parameters adaptively according to the current observation data, it is suitable for the situation of missing data or insufficient training data. At present, nonlinear adaptive prediction models mainly include two types, one based on series expansion and the other based on nonlinear function transformation. However, these two methods mainly focus on single-step prediction, and their multistep prediction needs further exploration [14, 15].

Based on the above analysis, the purpose of this paper is to develop a medium and long-term prediction model for avionics based on ESN, which requires sufficient accuracy of multistep prediction and as low computational complexity as possible. At the same time, due to the variation of avionics operating parameters, the statistical characteristics of the observed multivariate time series may change with time. Therefore, the model should have the function of updating parameters in real time and tracking the changing trend of faults online.

2. Related Work

Fault prediction is the core content of fault evaluation, which still belongs to fault diagnosis in a broad sense. It is an extended analysis of fault diagnosis and also the symbol of advanced PHM technology. The purpose is to predict the performance degradation trend and the remaining service life of the system components before the failure occurs. Trend prediction is to estimate the current state and future development trend of system components by statistical analysis of features. Residual life prediction can be regarded as an extended analysis of trend prediction. Usually, monitoring data are compared with characteristic historical trajectories so as to estimate the remaining normal service time of the previous system components. At the same time, there are methods based on experience and failure models. At present, the classification of fault prediction technologies has not been unified. By summarizing the research on mainstream fault prediction technologies, it can be divided into three categories, as shown in Figure 1. From the figure, we know that the knowledge-based approach has the widest application range, and the narrowest is the mechanism-based model. And the prediction accuracy and difficulty show the opposite trend.

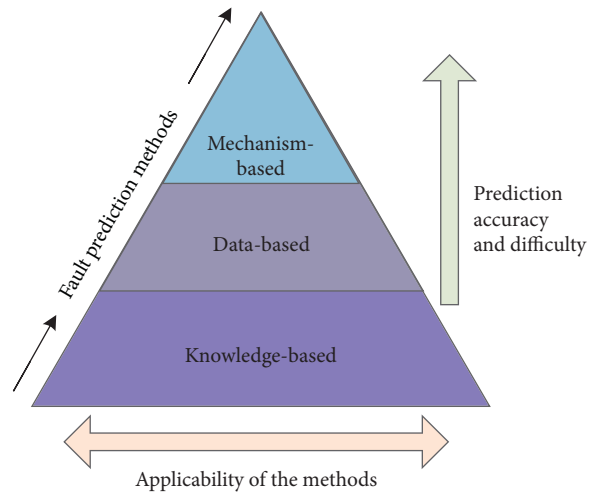


Figure 1: Classification of fault prediction methods.

Fault prediction method based on the mechanism model: This method requires that a mathematical model reflecting the physical law of performance degradation of the research object can be established. By integrating the environmental load measured by sensors with the damage model selected according to the failure mode of the system or component, precise prediction results can be calculated by calculating the component performance degradation caused by the accumulated load. The physical models commonly used to describe system or component failure include the crack propagation model, fatigue spalling propagation model, and so on. However, the failure physical model of the complex system is very complicated, and some physical characteristics are stochastic and complex, which limit the application of this method in practical engineering. Data-driven fault prediction methods [16, 17]: Through machine learning, multivariate statistical analysis, and other methods, the health behavior model of system components can be learned from historical monitoring data, and the future health status trend or remaining service life of the system components can be evaluated by means of trend change, threshold judgment, degradation curve comparison, and other means. The data-driven method is widely used in failure prediction of complex mechanical systems such as avionics because it does not need prior knowledge of system components and obtains key information from sensor historical data, reducing dependence on historical fault data. Fault prediction method based on an expert system: When the component or system lacks sensor monitoring data or the physical model of the system is difficult to establish, but there are enough historical failure data, the method can be used for data analysis, and the life distribution law of the object can be studied by means of fitting. The fault prediction method based on the life distribution model selects suitable life models such as exponential distribution, normal distribution, and Weibull distribution according to different object characteristics and then makes statistical analysis on historical data to determine the parameters of the distribution model. The fault tree analysis-based prediction method counts all the factors that may contribute to the occurrence of faults, establishes the logical block diagram, and realizes the determination and probability estimation of various possible causes of faults by analyzing the number of faults step by step. However, this method is only suitable for failure prediction of a large number of products, not specific to individual prediction and failure causes [18].

Events related to time can also be seen everywhere in real life, such as the click risk of buildings in thunderstorm weather, the number of WeChat steps a person takes every day or the amount of monthly payment, the annual number of business people, and the quarterly price index of a supermarket. Because this paper is based on the data time-stamp data to predict the failure time of avionics, it is a typical time series modeling problem. At present, the most commonly used models for chronological modeling are the autoregression (AR) model, moving average (MA) model, and auto regression roving average (ARMA) model. When the time series itself is not stationary series, if its increment is near zero, it can be called stationary series. However, the model has a conditional limit, which not only limits the highest order of the model but also limits the current random disturbance and the past sequence value. Later, relevant researchers started with features and tried to use machine learning methods to model. Machine learning is also popular in recent years [19, 20]. In terms of datasets and machine learning methods, it has improved a lot. It has evolved from simple at the beginning to complex and diversified now. From domestic and foreign scientific research achievements and a large number of experimental results, machine learning has been applied to all walks of life, such as the random forest model, neural network model, logistic regression model, naive Bayes model, and so on, and achieved good results. From the point of view of our country at present, with the enhancement of data acquisition ability, it has gradually evolved from general dataset to massive

dataset, and the computing hardware capacity of major Internet factories (namely, computing power) is also becoming more and more powerful. As a result, the model expression ability of the machine learning method gradually becomes stronger, and the learning ability of the dataset also becomes significantly stronger. However, a series of problems affecting prediction accuracy, such as long training time, gradient explosion, and easy over-fitting, should also be considered [21].

The analysis and prediction of the failure trend of avionics equipment can greatly reduce aircraft flight safety accidents and national economic losses. Research institutions and airlines are committed to the improvement and innovation of avionics fault prediction algorithms. To improve the stability and safety of avionics operation, it is imperative to predict and analyze the failure trend of avionics in time. Foreign research in the field of avionics fault prediction technology started early in 1950. The Palo Alto Company of the United States introduced the fault diagnosis device for the first time to maintain an aeroengine. In the 21st century, the corresponding intelligent methods have been paid attention to by many foreign researchers and widely used in various fields. In 2018, the avionics company Rolls-Royce partnered with an American AI company uptake to use artificial intelligence to predict aeroengine problems. Although neural network technology is widely used in fault prediction, many researchers often combine it with other technologies to solve the defects of neural network prediction, so as to achieve a better prediction effect [22]. Although the research in the field of avionics fault prediction started late in China, it has been devoted to the research of fault prediction technology. In the parameter prediction method, the traditional time series prediction method and artificial neural network intelligent method are used most. Domestic scholars often combine neural networks with other technologies to improve the prediction difficulty of electronic equipment such as complex systems. Generally speaking, people compare and screen out the prediction model with the best result from different prediction model methods and exclude the other models in order to improve the prediction accuracy. However, different prediction models are not exclusive; each has its own advantages and disadvantages. In the field of fault prediction, different models learn sequence information from different angles. If different models can be used to combine and complement each other, it is possible to further increase the accuracy of prediction. The mixed forecasting method improves the shortcomings of the single forecasting method, gives full play to the advantages of each method, and also makes up for their shortcomings. The fault prediction of an aeroengine can obtain the fault information of the engine in advance, so that the fault maintenance can be carried out in time, thus reducing the economic loss and improving safety.

The authors of [23] presented the case where the test mode of a single failure is sufficient to cover all multiple failures. Since the signed directed graph model can only make a qualitative analysis of the complex model for the electronic equipment class, the quantification of judgment accuracy is crucial. The introduction of a membership degree in fuzzy theory makes quantitative analysis possible and breaks the concept of either/or, but fuzzy theory ignores randomness when determining the boundary, so it is difficult for fuzzy theory to associate fuzziness with randomness. And the degree of membership is usually given by expert experience, which is inevitably subjective. The authors of the literature [24] first selected appropriate characteristic values for learning by the binary classification method, then inputted the selected key features into the fast model for learning time series information, predicted future time series by using the multicore method, and selected a large number of real stock history information for verification. The feasibility of the proposed time series prediction method in stock price prediction is illustrated. The authors of reference [25] revealed through the relevant data of bearings in the operation process of electronic equipment that the infinite hidden Markov model is applied to carry out health monitoring, and the hidden state of the single-layer model is divided, so as to achieve fault prediction. The ESN model is a promising multistep forward time series prediction strategy that has been used to predict time series data effectively. At present, the avionics system integrates airborne equipment parameter setting and instrument display function into a unique display control system through advanced liquid crystal display, integrated circuit, communication equipment and software technology, forming a centralized control, distributed management structure. Although the ESN model has good medium and long-term prediction performance, it has too many parameters to update and iterate.

As aircraft systems become more and more complicated, various organizations or individuals have developed a lot of analysis software to study the failure of avionics systems. Current avionics fault prediction systems are generally combined with artificial intelligence algorithms. Some excellent fault prediction methods based on the ESN model have appeared in some published journal papers. Although many reliable and efficient algorithms emerge endlessly, these algorithms are still difficult to deal with long-term fault prediction. It is still a long way to improve the performance of the ESN model to improve the fault prediction ability of avionics. Based on the above discussion, the main contributions of this paper are given as follows:

(1) This paper is the first time to apply the ESN model to the avionics forecasting field

(2) The method in this paper realizes the medium and long-term prediction of avionics. Compared with the short-term prediction, the research in this paper has more theoretical and practical significance.

3. ESN-Based Medium and Long-Term Avionics Fault Prediction

3.1. The ESN Model. The recursive neural network has a rich nonlinear dynamics mechanism, but the training process is mostly based on the gradient descent principle, the solving speed is not ideal, and there are local optimal

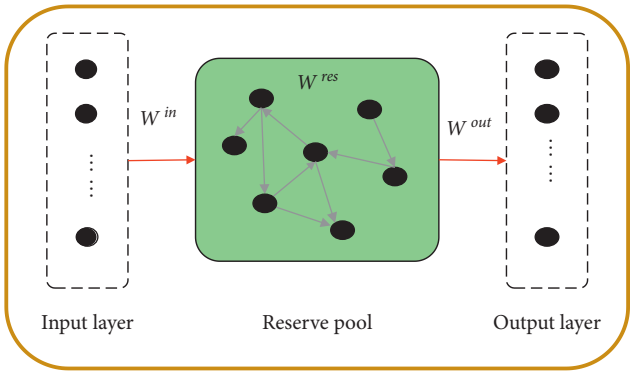


Figure 2: The typical structure of the ESN model.

problems. And the training process of recursive connection weights in the network is related to the output, so it is difficult to ensure the stability of the network. If the network is unstable, it cannot be predicted. Therefore, network stability is expected to be ensured in a relatively simple way. Network parameters may be given in advance before training and remain unchanged during training, or network stability can be ensured in real time through some adjustment mechanisms during training.

As a new type of recursive neural network, the traditional ESN includes input layer, hidden layer, and output layer, as shown in Figure 2. Among them, the hidden layer, also known as the reserve pool, is composed of large-scale nodes, which are sparsely connected randomly. In the ESN, the storage pool is used for information processing and storage. Before network training, network parameters are given in advance, and the input connection weights and the internal connection weights of the reserve pool are randomly initialized without changing. Only the connection weights between the reserve pool and the output layer need to be obtained through training.

At time t, state equation and network output of ESN are as follows:

$$\mathbf{S}(t) = f\left(\mathbf{W}^{\text{in}}\mathbf{I}(t) + \mathbf{W}^{\text{res}}\mathbf{S}(t-1)\right), \tag{1}$$

$$\mathbf{O}(t) = f^{\text{out}}\left(\mathbf{W}^{\text{out}}\mathbf{S}(t)\right), \tag{2}$$

where f represents the neuron activation function of the reserve pool. f^{out} represents the output unit activation function. Therefore, equations (1) and (2) can be rewritten as

$$\mathbf{S}(t) = \tanh\left(\mathbf{W}^{in}\mathbf{I}(t) + \mathbf{W}^{\text{res}}\mathbf{S}(t-1)\right), \tag{3}$$

$$\mathbf{O}(t) = \mathbf{W}^{\text{out}}\mathbf{S}(t). \tag{4}$$

Specifically, the ESN training process mainly includes the following steps. For a given learning task, the ESN parameters are properly initialized, including the input, reserve pool, and output node number. After the initialization, the ESN status is updated according to formula (3) under the input driver. To prepare for the subsequent calculation of the output weight, the state of the reserve pool at each moment is collected into a state matrix $\mathbf{Q}$, denoted as

$$\mathbf{Q} = \begin{bmatrix} s_1(1) & s_2(1) & \dots & s_N(1) \\ s_1(2) & s_2(2) & \dots & s_N(2) \\ \vdots & \vdots & \vdots & \vdots \\ s_1(l_{tr}) & s_2(l_{tr}) & \dots & s_N(l_{tr}) \end{bmatrix}_{l_{tr}\times N}, \tag{5}$$

where $\mathbf{S}(t) = [s_1(t)\ s_2(t)\ \dots\ s_N(t)]$ is the state of all neurons in the reserve pool at time t. Meanwhile, the expected signal corresponding to the input signal at each moment is denoted as

$$\mathbf{D} = \begin{bmatrix} d_1(1) & d_2(1) & \dots & d_L(1) \\ d_1(2) & d_2(2) & \dots & d_L(2) \\ \vdots & \vdots & \vdots & \vdots \\ d_1(l_{tr}) & d_2(l_{tr}) & \dots & d_L(l_{tr}) \end{bmatrix}_{l_{tr}\times L}. \tag{6}$$

The goal of the ESN training is to make the actual output $O(t)$ of the network approximate the expected value $D(t)$, i.e.,

$$\mathbf{d}(t) \approx \mathbf{O}(t) = \mathbf{W}^{\text{out}}\mathbf{S}(t). \tag{7}$$

Thus, the mean square error between $O(t)$ and $D(t)$ in the training stage is minimized.

$$MSE_{\text{train}} = \frac{1}{l_{tr}}\sum_{t=1}^{l_{tr}}\left(\mathbf{W}^{\text{out}}\mathbf{S}(t) - \mathbf{d}(t)\right). \tag{8}$$

The above problems are transformed into solving the least square method problem, namely,

$$\mathbf{W}^{\text{out}} = \underset{\mathbf{W}}{\text{argmin}}\,\|\mathbf{QW} - \mathbf{D}\|_2^2. \tag{9}$$

The solution to equation (9) can be found by using the following pseudoinverse algorithm:

$$\left(\mathbf{W}^{\text{out}}\right)^T = \left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{D}. \tag{10}$$

However, for high-dimensional states, the pseudoinverse algorithm will produce an inappropriate solution and overfitting phenomenon. The ridge regression training algorithm is used in this paper:

$$\mathbf{W}^{\text{out}} = \underset{\mathbf{W}}{\text{argmin}}\left\|\mathbf{QW} - \mathbf{D}\right\|_2^2 + \gamma\left\|\mathbf{W}\right\|_2^2. \tag{11}$$

The solution to equation (11) is

$$\left(\mathbf{W}^{\text{out}}\right)^T = \left(\mathbf{Q}^T\mathbf{Q} + \gamma\mathbf{E}\right)^{-1}\mathbf{Q}^T\mathbf{D}, \tag{12}$$

where γ is the regularization coefficient, and E is the identity matrix with dimension N.

ESN is a new type of the burst neural network which is widely used in nonlinear time series prediction modeling. Only the output weights need to be solved by applying the linear regression method, which greatly reduces the complexity of system modeling. However, in practical application, it is found that for a large-scale reserve pool, the small change of system state variable may lead to a huge change in output weight, resulting in an ill-conditioned solution. At present, the ESN model is not applied to long-term fault prediction of

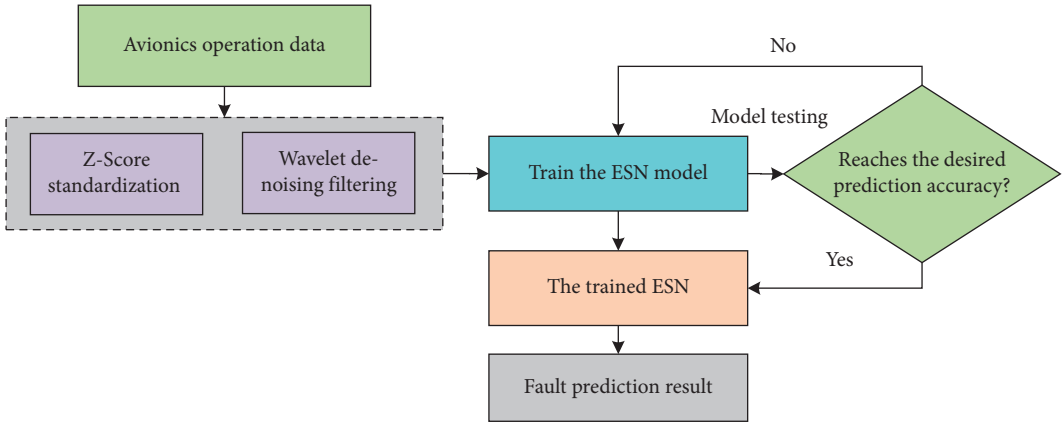


Figure 3: The framework of ESN-based avionics medium and long-term fault prediction.

avionics. Based on the above discussion, the ESN model and its application in long-term fault prediction of avionics are shown in Figure 3. It mainly includes the data preprocessing, model training and testing, and fault prediction.

4. Experimental Results and Analysis

4.1. Experimental Data Introduction. In this paper, the data collected by sensors and other devices on real avionics are collected from 12 different avionics devices, and the data collected by each device ranges from thousands to tens of thousands. The avionics collected contain 28 characteristic values: equipment number, time, temperature, voltage, and so on. In the ESN in this paper, if the equipment is of the same type or has similarities, the data can be integrated together as a training set to predict the possible failure time interval of such equipment.

This paper uses Matlab and Python to process avionics data. Python has a variety of functions, not only for software and front-end development but also for powerful functions in mathematical calculation. Python even has a special open-source library for the development of machine learning deep learning, which has been hot in recent years. PyTorch is an open-source learning library for deep learning of machine learning in Python. This module can greatly improve the efficiency of deep network calculation and training, not only improving the accuracy of training in the process of deep network calculation but also providing a guarantee for the construction of a more complex deep network model in the future.

4.2. Experimental Results Analysis. In order to verify the influence of different sample ratios on the ESN algorithm, this paper found in the experiment that the accuracy and stability of the algorithm could be improved if some process data before and after test time points were added as unmarked samples to participate in training. What is uncertain is how many recorded samples (HS), pretest samples (BS), and posttest samples (AS) can improve accuracy. In this paper, the change of accuracy was tested by constantly adjusting the proportion of the three in the unmarked sample. As shown in Figure 4, the X-axis represents the proportion of HS and AS in the unmarked sample, and the Y-axis represents the proportion of HS and BS.

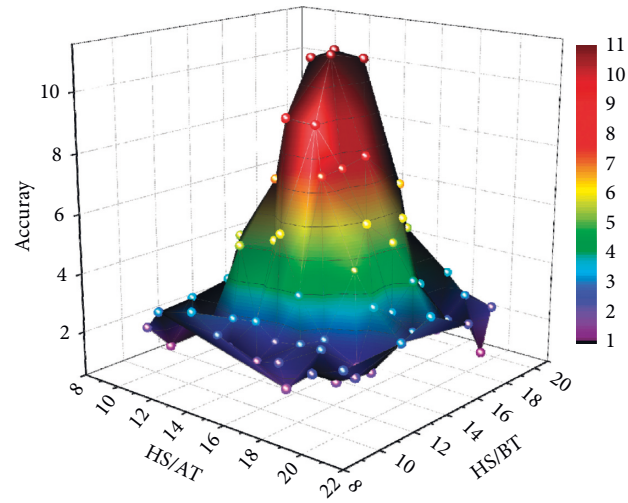


Figure 4: The effects of sample size on accuracy before and after testing.

It can be seen that the added test node sample effectively improves the accuracy and can reach the maximum value. This is because the process data before and after the addition of test nodes in the unmarked sample can improve the features of failure prediction during training. Since it is difficult to extract precise fault signal features under complex noise, the training of unlabeled samples is equivalent to feature extraction again, which improves the prediction accuracy.

The influence of the proportion of different training samples on the prediction accuracy of the ESN model should be further demonstrated. It can be seen from Figure 5 that with the increase of the proportion of the training set, the distribution of prediction error is concentrated towards zero, and the box graph is more compact, proving that the prediction accuracy becomes higher. When the training proportion is 80%, the prediction accuracy of the model reaches the highest, but when the proportion of the training set continues to increase (set as 90%), the prediction accuracy of the model shows a trend of decline, possibly due to the phenomenon of overfitting of the model.

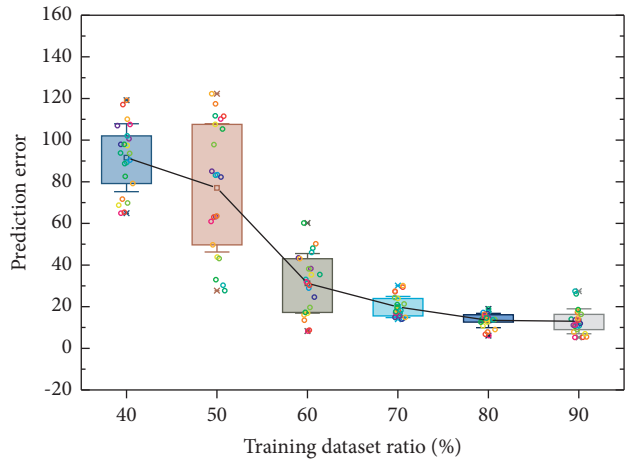


Figure 5: The prediction accuracy of ESN under different training dataset ratios.

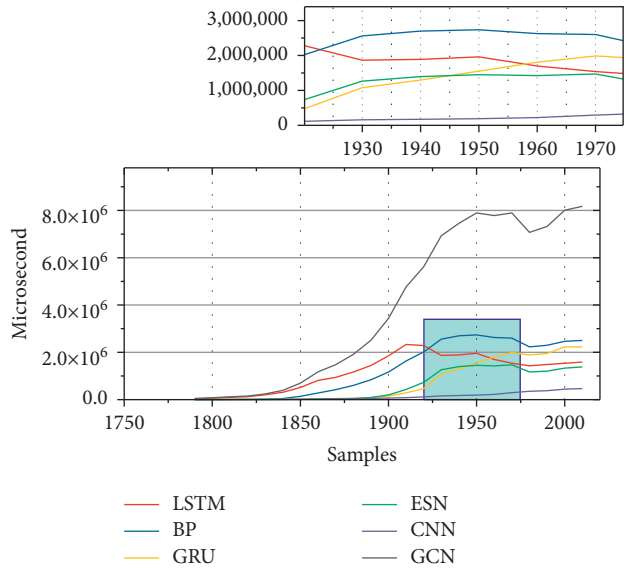


Figure 6: Prediction time of different methods with increasing sample size.

In addition to the prediction accuracy, the running time of the model is also an important factor to be considered. Figure 6 gives the prediction time of different methods (LSTM (long short-term memory), BP, GRU (gated recurrent unit), ESN, CNNs (convolutional neural networks), and GCN (graph convolutional network)) with increasing sample size. As can be seen from the figure, CNN and GCN models have the longest running time due to their deep network structure. Although they may have high predictive accuracy, their long training time is one of the major drawbacks. In contrast, the running time of the remaining four models will be much reduced. Among them, although the running time of the ESN model proposed in this paper is not the lowest, it is completely acceptable for the current computer level. It can be seen from the above results that the method proposed in this paper not only has high prediction accuracy but also has a relatively short running time, so the comprehensive performance of the model is quite good.

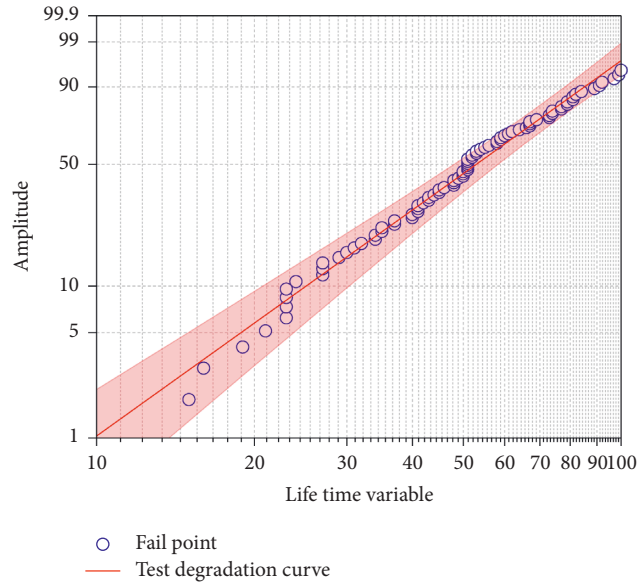


Figure 7: Fault degradation trajectory graph under 90% training dataset.

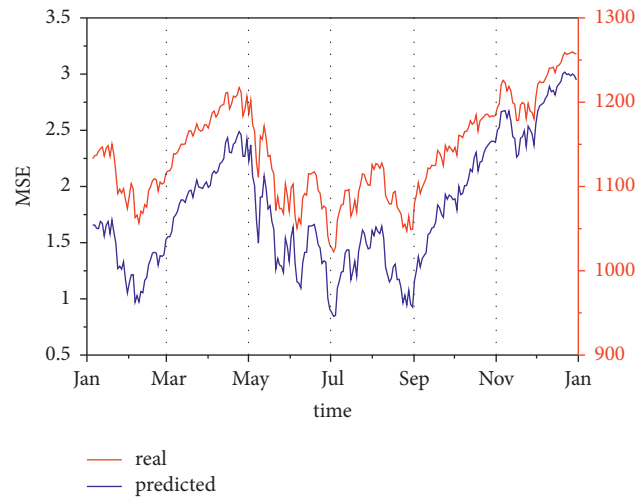


Figure 8: Medium and long-term fault prediction results of the ESN model.

In order to eliminate the influence of data selection on comparison results, the ESN model is used to process the training and test data divided in the previous paper in the same way, and the test degradation curve based on ESN is obtained. It can be clearly observed from Figure 7 that as the operation cycle of test data increases, the smaller the fault prediction interval becomes, the more concentrated it becomes. From the results, its prediction accuracy will be higher, and it is reasonable to establish a confidence interval with 90% test data. It shows that the proposed method can well track the fault development process of avionics and has good performance in fault prediction.

In order to further demonstrate the medium and long-term performance of the proposed method, Figure 8 shows the failure prediction results of the ESN model for avionics

in one year. It can be seen from the figure that the MSN value of the model is relatively large during the five months from January to May, mainly because the model did not fit the process data well at the beginning, so the prediction accuracy is not very good. With the increase of time and the accumulation of data, the predicted values of the proposed ESN model can track the real fault data well from June to January of next year and obtain a low MSE value, which indicates that the proposed method has a good performance of medium and long-term fault prediction.

5. Conclusions

With the rapid development of the modern aviation industry, the degree of automation and intelligence of avionics has been significantly improved. The normal operation of avionics depends on the close cooperation between various systems. However, the failure performance of avionics is also very complicated due to the complex and bad aviation environment and heavy equipment workload. In view of the equipment with abnormal operation status, the failure prediction is carried out by using the operation data of the avionics system.

In this context and in view of the existing research on the failure of avionics equipment for the long-term prediction of the problem, specifically, the preprocessed data are input into the model for training and testing, and the ideal medium and long-term fault prediction results are finally obtained. Finally, experimental results verify the superiority and effectiveness of the proposed method. Although the method in this paper has achieved a good prediction effect, ESN is still a shallow model. When encountering big data, the improved deep learning model will be worth studying.

Data Availability

The datasets used during the current study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Lim, A. Gardi, R. Sabatini et al., "Avionics human-machine interfaces and interactions for manned and unmanned aircraft," *Progress in Aerospace Sciences*, vol. 102, pp. 1–46, 2018.

[2] Z. H. U. Shaofan, T. Jian, and J. Ml. Gauthier, "A formal approach using SysML for capturing functional requirements in avionics domain," *Chinese Journal of Aeronautics*, vol. 32, no. 12, pp. 2717–2726, 2019.

[3] Q. Liang, X. Zhang, and Q. Hu, "The reliability prediction of torpedo electronic components in service life," *Journal of Navigation*, vol. 74, no. 6, pp. 1367–1380, 2021.

[4] A. Abuelnaga, M. Narimani, and A. S. Bahman, "Power electronic converter reliability and prognosis review focusing on power switch module failures," *Journal of Power Electronics*, vol. 21, no. 6, pp. 865–880, 2021.

[5] J. Ni, L. Han, J. Pan et al., "Evolution of contact performance of industry electrical connector based on reliability accelerated testing," *Advances in Mechanical Engineering*, vol. 13, no. 2, Article ID 168781402199882, 2021.

[6] J. Walther and M. Weigold, "A systematic review on predicting and forecasting the electrical energy consumption in the manufacturing industry," *Energies*, vol. 14, no. 4, p. 968, 2021.

[7] A. R. Nia, A. Awasthi, and N. Bhuiyan, "Industry 4.0 and demand forecasting of the energy supply chain: a literature review," *Computers & Industrial Engineering*, vol. 154, Article ID 107128, 2021.

[8] M. Ortega, E. Ivorra, A. M. Juan, P. Venegas, J. Martinez, and M. Alcaniz, "MANTRA: an effective system based on augmented reality and infrared thermography for industrial maintenance," *Applied Sciences*, vol. 11, no. 1, p. 385, 2021.

[9] T. Andersen and B. Jæger, "Circularity for electric and electronic equipment (EEE), the edge and distributed ledger (edge&DL) model," *Sustainability*, vol. 13, no. 17, p. 9924, 2021.

[10] C. Wang, T. Ji, and F. Mao, "Prognostics and health management system for electric vehicles with a hierarchy fusion framework: concepts, architectures, and methods," *Advances in Civil Engineering*, vol. 2, pp. 185–199, 2021.

[11] C. S. Kulkarni and K. Goebel, "Joint special issue on PHM for aerospace systems," *International Journal of Prognostics and Health Management*, vol. 12, no. 3, pp. 25–38, 2021.

[12] P. M. Congedo, C. Baglivo, A. K. Seyhan, and R Marchetti, "Worldwide dynamic predictive analysis of building performance under long-term climate change conditions," *Journal of Building Engineering*, vol. 42, Article ID 103057, 2021.

[13] I. A. Alobaidi, J. L. Leopold, and A. A. Allami, "Predictive analysis of real-time strategy games: a graph mining approach," *Wiley Interdisciplinary Review: Data Mining and Knowledge Discovery*, vol. 11, no. 2, p. e1398, 2021.

[14] G. Shi, C. Qin, J. Tao, and C Liu, "A VMD-EWT-LSTM-based multi-step prediction approach for shield tunneling machine cutterhead torque," *Knowledge-Based Systems*, vol. 228, Article ID 107213, 2021.

[15] R. Chandra, S. Goyal, and R. Gupta, "Evaluation of deep learning models for multi-step ahead time series prediction," *IEEE Access*, vol. 9, Article ID 83105, 2021.

[16] L. Wen, X. Li, L. Gao, and Y Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[17] H. Chen, B. Jiang, and S. X. Ding, "Data-driven fault diagnosis for traction systems in high-speed trains: a survey, challenges, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 12, p. 5546, 2020.

[18] N. Hou, N. Ding, and S. Qu, "Failure modes, mechanisms and causes of shafts in mechanical equipment," *Engineering Failure Analysis*, vol. 136, Article ID 106216, 2022.

[19] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: a survey and outlook," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–36, 2022.

[20] J. Grimmer, M. E. Roberts, and B. M. Stewart, "Machine learning for social science: an agnostic approach," *Annual Review of Political Science*, vol. 24, no. 1, pp. 395–419, 2021.

[21] S. Zhang, N. Qu, T. Zheng, and C Hu, "Series arc fault detection based on wavelet compression reconstruction data enhancement and deep residual network," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.

[22] S. Lu, Q. Zhang, G. Chen, and D Seng, "A combined method for short-term traffic flow prediction based on recurrent

neural network," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 87–94, 2021.

[23] R. Hemmati, H. Mehrjerdi, and S. M. Nosratabadi, "Resilience-oriented adaptable microgrid formation in integrated electricity-gas system with deployment of multiple energy hubs," *Sustainable Cities and Society*, vol. 71, Article ID 102946, 2021.

[24] Y. Hu, P. Zhan, Y. Xu, J. Zhao, Y. Li, and X Li, "Temporal representation learning for time series classification," *Neural Computing & Applications*, vol. 33, no. 8, pp. 3169–3182, 2021.

[25] L. Piho and M. Kruusmaa, "Subsurface flow path modeling from inertial measurement unit sensor data using infinite hidden Markov models," *IEEE Sensors Journal*, vol. 22, no. 1, pp. 621–630, 2022.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 5408470, 12 pages
https://doi.org/10.1155/2022/5408470

Research Article

Attribute-Based Policy Evaluation Using Constraints Specification Language and Conflict Detections

Wei Sun

School of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China

Correspondence should be addressed to Wei Sun; sunny810715@xynu.edu.cn

Received 29 June 2022; Accepted 22 August 2022; Published 5 September 2022

Academic Editor: Xingsi Xue

Attribute-based access control (ABAC) has attracted widespread interest and has become an ideal mechanism due to its flexibility characteristic and the powerful expressiveness for various security policies, such as the separation-of-duty constraint and cardinality constraint. The formulation of appropriate ABAC policies is critical for ensuring system security and robustness. However, conflicts occur frequently in existing state-of-the-art systems. Most conventional detection methods either lack the evaluation of the policy quality or consider no constraint. To resolve these problems, a novel method for the ABAC policy evaluation is proposed in this study. First, to meet diverse organizational requirements, we use the attribute-based constraints specification language to uniformly formulate and specify the conflict relations among attributes and present the satisfiability of conflict relations. Second, to comprehensively detect the conflict problems, we present the evaluation criteria for conflicts on attributes and rules and propose a novel algorithm for detecting conflicts. Last, we validate the effectiveness and efficiency of the proposal through experiments, which demonstrate that it not only improves the policy quality but also reduces the conflicting number and conflicting probability.

1. Introduction

With the high-speed development in high-performance computing and mobile-information technology, security has been considered as a fundamental requirement for the research fields such as the Internet of Things (IoT), smart contracts, blockchains, and the industrial information integration system [1]. There are large amounts of data storage and resource sharing in distributed and collaborative environments, and enterprises need to employ some means to ensure the integrity and confidentiality of information systems. As the main benchmark, the role-based access control (RBAC) model had been widely used for system implementation and management over the last few decades [2]. However, it is identity dependent and lacks flexibility and extendibility. As an alternative, attributes are employed to describe the features of entities. The attribute-based access control (ABAC) overcomes the limitations of RBAC, captures fine-grained access requirements, and becomes very attractive, particularly for large-scale distributed and collaborative systems [3]. It has gained much attention in both academia and industry [4] in recent years.

Actually, the ABAC policy rule is the combinational form of different attribute value pairs of subjects, objects, environments, and operations. The policy engineering [5, 6] is to find a suitable ABAC rule set, which is regarded as the most important step for implementing the ABAC mechanism. Xu and Stoller [7] were the first to study the ABAC policy mining problem from the given access control lists or matrices and proposed a bottom-up resolution (represented as the Xu-Stoller for simplicity). To reduce the scale of ABAC rules, Das et al. [8] used the Gini impurity and presented a policy mining method. Das et al. [9] also presented a visual method, called VisMAP, which mined ABAC policies based on a given authorization list.

The ABAC policy is very flexible and extendable. However, values of the attribute-expression conditions specified by different rules are partially identical. If the access decisions of the rules that have identical attribute values are inconsistent, then there exist conflicts among such rules, and

policy maintenance becomes difficult [10]. Therefore, how to detect the conflicts among policies has become an urgent problem to be resolved. Royer and De Oliveira [11] separated the existing conflict detection mechanisms for the eXtensible Access Control Markup Language (XACML) into three different types. The dynamic and testing detections depend on access requests, while the static detection was based on the rule set without generating requests. Jabal et al. [12] summarized the related research on static conflict detections and involved five different variants. St-Martin and Felty [13] converted the XACML policies into the Coq code that included the effect type and the SRAC that represented the "subject-resource-action-condition." Rezvani et al. [14] proposed a method to translate an XACML policy into the form of ASP programming, verified the properties of the policy by an analysis tool, and then, validated its effectiveness. Zheng and Xiao [15] visually specified ABAC rules, converted them into a set of binary sequences and presented a novel method for conflict detection. Shu et al. [16] proposed an optimized method to detect explicit conflicting rules from the given ABAC policy, which utilized the method of rule reduction to eliminate the redundancy of the rules and then adopted the method of binary search to improve the efficiency of the detections. To extend the detecting scope while improving the efficiency of the detections, based on the intersection of rule pairs, Liu et al. [17] proposed a novel approach for detecting both explicit and implicit conflicts. However, the existing methods do not consider the constraint requirements during the detecting processes.

To comprehensively capture the different organizational requirements while ensuring the ABAC system security and confidentiality, an important feature of the ABAC mechanism is to be able to specify and perform the cardinality constraint and the separation-of-duty constraint (SOD). These constraint policies are not relevant to the specific access control mechanism [18]. To implement several classical access control models, Jin et al. [19] presented a novel framework, called $ABAC_{\alpha}$, which specified constraints on the attribute-assignment relationships using the policy specification language. However, the constraints in ABACα were dependent on the specific events, and they became ineffective if the attribute assignments varied. To address this problem while uniformly specifying various types of ABAC constraints, Bijon et al. [20] proposed the attribute-based constraint specification language (ABCL). Based on the subject similarity, Helil and Rahman [21] used the ABAC constraint to check and determine the potential relationships between different entities. To further specify and verify the SoD constraints in ABAC systems, Jha et al. [22] presented a novel approach for analyzing the complexity of enforcing the SOD constraints. To verify whether a set of users could be replaced by another user set, Roy et al. [18] presented the employee-replacement problem with multiple constraints and then provided a scheme for solving the problem. Furthermore, to automatically derive ABAC rules from the conventional access control documents, Alohaly et al. [23] proposed a framework for policy extraction using natural language processing techniques. However, most conventional methods only focus on specifying or formulating constraints on attributes, which do not consider the influence of constraints on the ABAC rules and lack the evaluation of the policy quality. Thus, conflict problems among ABAC rules arise frequently while using the existing research methods.

To resolve the abovementioned problems, this study proposes the attribute-based policy evaluation using constraints specification language and conflict detections (ABPE_CSL&CD). To sum up, the main contributions of this work are as follows:

(1) To flexibly suit organizational requirements, while ensuring system security, we use the ABCL to uniformly formulate and specify the conflict relations among attributes and propose the satisfiability of conflict relations. We take the reconstructed ratio as the evaluation criterion and demonstrate the efficiency of the ABPE_CSL&CD using real datasets.

(2) To comprehensively detect the ABAC conflict problems, while ensuring the system robustness, we present the classification representations for conflicting rules and propose a novel method for conflict detection. We take the conflicting number and conflicting probability as the evaluation criteria and demonstrate the effectiveness of the ABPE_CSL&CD using synthetic datasets.

The rest of the article is organized as follows: Section 2 introduces some necessary preliminaries. Section 3 proposes a novel policy evaluation method and presents an algorithm for conflict detection. We present the experimental analysis in Section 4 and conclude the article and discuss future works in Section 5.

2. Preliminaries

In this section, some preliminaries are presented, including the basic components of the ABAC, basic components of the ABCL, and conflict problems in the ABAC.

2.1. Basic Components of ABAC. According to the ABACα [19], the ABAC model mainly consists of the following sets, relations, and functions:

(1) Sets S, O, and E represent all the subjects, objects, and environments in which the access control occurs, respectively. Set OP represents all the operations that are permitted or denied to be performed on the object resources. Sets SA, OA, and EA, respectively, represent the identifier names of the subjects, objects, and environments, which can be categorized into multi-valued and single-valued types. For instance, the *role* attribute is multi-valued as an employee may own more than one role in an organization, while the *id* attribute is single-valued as any employee in the organization has a unique identifier value.

(2) Functions $atttype(att)$ and $range(att)$, respectively, represent the type and value domain for a specific entity attribute att, which can be formalized as

$$\forall \text{att} \in SA \cup OA \cup EA\text{: atttype(att)} \in \{\text{atomic, set}\}, \text{range(att)} = \left\{val_i^{\text{att}} | i \in Z^+\right\}. \quad (1)$$

For the sake of convenience, the environmental elements E and EA are not taken into account here.

(3) Relation $SSAV$ represents many-to-many assignments relationship of subjects and their attribute-expression conditions, for any user attribute att, which can be formalized as follows:

$$\forall \text{att} \in SA, SSAV\text{: } S \longrightarrow \begin{cases} \text{range(att)}, \text{if atttype(att)} = \text{atomic}, \\ 2^{\text{range(att)}}, \text{if atttype}(att) = \text{set}. \end{cases} \quad (2)$$

Similarly, $OOAV$ can be formalized as

$$\forall \text{att} \in OA, OOAV\text{: } O \longrightarrow \begin{cases} \text{range(att)}, \text{if atttype(att)} = \text{atomic}, \\ 2^{\text{range(att)}}, \text{if atttype(att)} = \text{set}. \end{cases} \quad (3)$$

Furthermore, the value set of attribute att assigned to any user or object entity en is denoted as function $val(en, att)$.

(4) Function $Assigned\_Entities_{EN,att}$ returns the entities with respect to a specific attribute value $attval$, which can be formalized as follows:

$$\forall att \in (SA \quad or \quad OA), \exists \text{attval} \in \text{range(att)}: \text{Assigned\_Entities}_{S\cup O,\text{att}}(\text{attval}) = \{en | \exists en \in (S \quad or \quad O), (\text{val}(en, \text{att}) = \text{attval} \wedge \text{atttype(att)} = \text{atomic}) \quad or (\text{attval} \in \text{val}(en, \text{att}) \wedge \text{atttype(att)} = \text{set})\}. \quad (4)$$

2.2. Basic Components of ABCL. The key component of the ABCL [20] is conflict relations, which are used to determine whether the policy conditions, such as mutually exclusive constraints or cardinality constraints, can be satisfied. Attribute-based conflicts can occur in several ways, in which two critical conflicting variants are considered:

(1) The single-attribute conflict is only applicable for the multi-valued attributes and can be formally expressed in formulation (5). In the formulation, set *Single_Conf_Set* contains various types of the constraint requirements, such as mutual exclusions cardinality constraints, precondition constraints, and so on. Each element of the *Single_Conf_Set* is a 2-tuple form, denoted as (*attval*, *limit*), where *attval* represents a set of conflicts existing in the single-attribute values, and *limit* represents the threshold value for satisfying the security constraint.

$$\forall \text{att} \in (SA \quad or \quad OA), \text{atttype(att)} = set: \text{Single\_Conf\_Set}_{S,O,\text{att}} = \{\text{avset}_1, \text{avset}_2, \ldots, \text{avset}_n\}, \text{where} \quad \text{avset}_i(att) = (\text{attval}, \text{limit}), \text{attval} \in 2^{\text{range(att)}}, and \quad 1 \le \text{limit} \le |\text{attval}|. \quad (5)$$

(2) The cross-attribute conflict is applicable for both the single-valued and multi-valued attributes and can be formally expressed in formulation (6). In the formulation, two different attribute sets are involved, which are represented as *Aattset* and *Rattset*, respectively. The values of one attribute in *Aattset* restrict those of the other one in *Rattset*. Further, set *Cross_Conf_Set* also contains the constraint specifications for different attributes as shown in formulation (6), where each element is a function, named $attfun_i$.

$$\forall (\text{Aattset, Rattset}) \subseteq (SA \quad or \quad OA), \forall att \in (\text{Aattset} \quad or \quad \text{Rattset}): Cross\_Conf\_Set_{S,O,\text{Aattset,Rattset}} = \{\text{attfun}_1, \text{attfun}_2, \ldots, \text{attfun}_n\}, where \quad attfun_i(att) = (\text{attval}, \text{limit}), 1 \le \text{limit} \le |attval|, \left(\text{attval} \in 2^{\text{range(att)}}, \text{atttype}(\text{att}) = \text{set}\right) or (\text{attval} \subseteq \text{range(att)}, \text{atttype}(\text{att}) = \text{atomic}). \quad (6)$$

The ABCL also presents two nondeterministic functions, written as $OE(X)$ and $AO(X)$. $OE(X)$ selects one element from the set X, while $AO(X)$ returns the other elements from X except for the element with $OE(X)$. Both of them are related to contexts. This is because there exists an equation $\{OE(X)\} \cup AO(X) = X$ for the given set X.

2.3. Conflict Problems in ABAC. Taking the ABAC policy specified by the XACML [11] as an example, there exist conflict problems among the rules. Several concepts such as attribute expression, policy rule, access request, and conflict are taken into account in this study, which are described as follows:

(1) Attribute expression *ap*: it can be formalized as a triple $ap = (\text{attribute identifier, operator, and attribute values})$, where the *operator* can take different comparison operators.

(2) Policy rule *ar*: it can be represented as a quadruple $ar_i = (S_i^{ap}, O_i^{ap}, op_i, d_i)$, where op_i represents the operating action on the object resources, d_i represents the positive or negative access decision, which takes only two possible values: *permitted* or *denied*. S_i^{ap} or O_i^{ap} is represented as the set of combinations of various attribute-expression conditions for the subjects or objects. The set of n different rules $ar_1, ar_2, \cdots, ar_n$ constructs an ABAC policy, denoted as $P = \{ar_1, ar_2, \ldots, ar_n\}$.

(3) Access request *req:* it can be formalized as a triple $\text{req} = (S_{\text{req}}^{ap}, O_{\text{req}}^{ap}, op_{\text{req}})$, which indicates that the subjects of S_{req}^{ap} request to perform the operation op_{req} on the objects of O_{req}^{ap}.

(4) Conflict: for the given rules *ari* and *arj,* which satisfy a single request *req* simultaneously while owning the identical attribute identifiers, if the intersections of the attribute-expression conditions of the rules have common attribute values, the operation sets overlap, but if the access decisions are distinct, then there exists a conflict between *ari* and *arj.*

3. Methodology

The proposed ABPE_CSL&CD is threefold: (1) the formulation and specification for conflict relations among attributes, (2) classification representations of conflicting rules, and (3) ABAC conflict detections. Specifically, based on the conventional policy-engineering method, we first construct an initial policy set that takes no conflicts into consideration. Subsequently, according to the requirement descriptions of usage scenarios, such as the banking businesses, we utilize the ABCL to formulate and specify the conflict relations in a single attribute and multiple attributes. Meanwhile, we categorize the conflicting rules into two classifications and present the evaluation criteria for conflicting rules and attribute-conflict relations. Last, we present a novel algorithm for detecting conflicts from the initial policy set and evaluate the performance of the proposal through experiments. The framework of the ABPE_CSL&CD is presented in Figure 1.

3.1. Formulation and Specification for Conflict Relations among Attributes. In this section, an extensive case study in the banking-domain scenario is presented, which utilizes the standard ABCL to formulate various conflict relations among attributes, in order to express the business requirements and ensure system security.

3.1.1. Requirement Descriptions for the Banking Businesses. First, the corresponding attribute characteristics of subjects and objects are presented in Tables 1 and 2, respectively. Each subject is a user who is assigned attributes *id* and *style*. The attribute *role* represents the job responsibility. *trustness_level* and *work_year* are the other two attributes of the subjects, of which the values can be denoted using the comparison operation expressions, such as $trustness\_level \geq 11$, and $work\_year \geq 8$. The descriptions of object features are omitted owing to the space limitation.

According to Tables 1 and 2, the organizational requirements with various constraints are stated as follows:

Const 1: any user can obtain 5 *benefit businesses* at most

Const 2: any user cannot have both the *cashier* and *accountant* roles

Const 3: any user cannot obtain both the bf_1 and bf_2*benefit businesses*

Const 4: any user can together obtain 5 *loan* and *card businesses* at most

Const 5: if the *trustness_level* value of a user is less than 5, then this user cannot obtain more than 1 benefit from $\{bf_1, bf_2, and\ bf_3\}$

Const 6: the number of users obtaining the *house loan business* is no more than 12

Const 7: any two users cannot own the same *id* value

Const 8: if a user has the *house* and *car loan businesse*s and also has more than one *card business*, then this user cannot obtain any *benefit business*

3.1.2. ABCL Specifications for the Security Requirements. Next, the standard ABCL is used to specify the security requirements mentioned above, including the declarations and initializations of different conflict relations in a single attribute and multiple attributes, in which *SMERole* represents a related set of mutually exclusive constraints towards the *role* attribute values, and *OMEBenefit* is another conflict set towards the *benefit business* attribute. Similarly, *SOMETB* and *SOMETWB* represent mutually exclusive conflict sets of the *trustness_level* with *benefit business* and the *trustness_level* and *work_year* with *benefit business* by the *Cross_Conf_Set* variant, respectively. Here, assume that the number of attributes, which are assigned to an entity for satisfying each constraint relation from *Single _Conf_Set* or *Cross_Conf_Set* is less than the threshold *limit*.

(1) Declaration and initialization for *Single_Conf_Set*

(a) $\text{Single\_Conf\_Set}_{S,\ \text{role}}$ SMERole:SMERole = $\{\text{avset}_1\}\text{avset}_1(\text{role}) = (\{\text{cashier, accountant }\}, 2)$

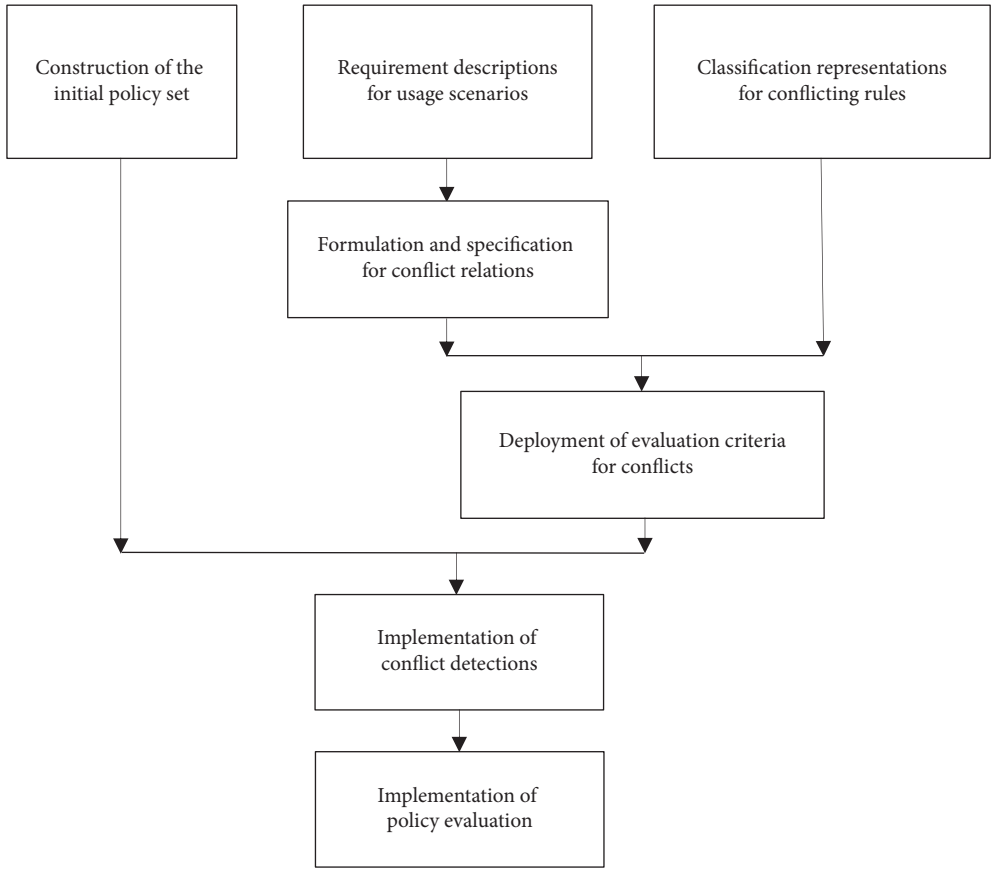


Figure 1: The framework of the ABPE_CSL&CD.

Table 1: Attribute characteristics of subjects.

SA	Type of *SA*	Range of *SA*
Id	*atomic*	$\{id_1, id_2, \ldots id_{20}\}$
Style	*atomic*	{*client and banking employee*}
Role	*set*	{*customer, general employee, cashier, accountant, junior, assistant manager, and manager*}
trustness_level	*atomic*	{1, 2, . . ., 15}
work_year	*atomic*	{1, 2, . . ., 30}

Table 2: Attribute characteristics of objects.

OA	Type of *OA*	Range of *OA*
General business	*atomic*	{*deposit and withdrawal*}
Loan business	*set*	{*house, car, and education*}
Benefit business	*set*	$\{bf_1, bf_2, \ldots, bf_{10}\}$
Card business	*set*	$\{Card_1, card_2, \ldots, card_{12}\}$

(b) Single_Conf_Set$_{O,\ \text{benefit business}}$ OMEBenefit: OMEBenefit = {avset$_1$}avset$_1$(benefit business)=({bf$_1$, bf$_2$}, 2)

(2) Declaration and initialization for *Cross_Conf_Set*

(a) Cross_Conf_Set$_{S,\ O,\ \text{trustness_level},\ \text{benefit business}}$ SOMETB:SOMETB = {attfun$_1$, attfun$_2$} attfun$_1$(trustness_level)=("≤5", 2), attfun$_1$(benefit business)=({bf$_1$, bf$_2$, bf$_3$}, 2) attfun$_2$(trustness_level)=("≤5", 1), attfun$_2$(benefit business)=({bf$_1$, bf$_2$, bf$_3$, bf$_4$, bf$_5$}, 1)

(b) Cross_Conf_Set$_{O,\ \{\text{loan business, ccard business}\},\ \text{benefit business}}$ OMELCB:OMELCB = {attfun1}attfun1(loan business)=({house, car}, 2), attfun1(ccard business)=({card1, card2, . . ., card12}, 2), attfun$_1$(benefit business)=({bf$_1$, bf$_2$, . . ., bf$_{10}$}, 1)

Then, the conflict relations and functions are used to express the security requirements in the ABCL form as follows:

Spec 1: $|\text{benefit\_business}(OE(S))| \leq 5$

Spec 2: $|\text{role}(OE(S)) \cap OE(\text{SMERole}).\text{avset}| < OE(SMERole)$. limit.

Spec 3: $|\text{benefit\_business}(OE(S)) \cap O\ E(OME\text{Benefit}).\text{avset}| < OE(\text{OMEBenefit}).\text{limit}$.

Spec 4: $|\text{loan\_business}(OE(S)) \cap ccar\,d\_business(OE(S))| \leq 5$.

Spec 5: $|\text{trustness\_level}(OE(S)) \cap OE(\text{SOMETB})(\text{trustness\_level}).\text{avset}| < OE(SOMETB)(trustness\_level).limit \Rightarrow |bene\,f\,it\_business(OE(S)) \cap OE(\text{SOMETB})(\text{benefitbusiness}).\text{avset}| < OE(SOMETB)(\text{benefitbusiness}).limit$.

Spec 6: $|\text{Assigned\_Entities}_{S\cup O,\text{loan business}}(\text{house})| \leq 12$.

Spec 7: $id(OE(S)) \neq id(OE(AO(S)))$.

Spec 8: $(|\text{loan\_business}(OE(S)) \cap OE(\text{OMELCB})(\text{loanbusiness}).\text{avset}| \geq OE((OMELCB)(\text{loanbusiness}).limit) \wedge (|\text{ccard\_business}(OE(S)) \cap OE(\text{OMELCB})(\text{ccardbusiness}).\text{avset}| \geq OE(OMELCB)(\text{ccardbusiness}).limit) \Rightarrow |\text{benefit\_business}(OE(S)) \cap OE(\text{OMELCB})(\text{benefitbusiness}).\text{avset}| < OE(OMELCB)(\text{benefitbusiness}).\text{limit}$.

3.2. Classification Representations of Conflicting Rules. According to the description of conflicts, if multiple rules satisfy the same access request, while their access decisions are different, then a conflict occurs among these rules. The conflicting rules can be directly compared and are easy to be detected by implementing the static analysis or using the existing tool before the system runs. They are referred to as explicit conflicting rules and are commonly seen in the policy set, such as ar_5 and ar_6, as shown in Table 3. Furthermore, other rules that seem to be not directly comparable, which are called implicit ones, and they still exist. For instance, implicit conflicts occur between ar_2 and ar_4 when the access request is ({*role* = {*general employee*}, *trustness_level* > 12}, {*loan business* = {*house*}, *card business* = {$card_2$}}, *apply for*). Similarly, ar_3 and ar_5 become implicit conflicting rules for a given request such as ({*role* = {*general employee*}, *trustness_level* > 8, 10 < *work_year* < 20}, {*benefit business* = {bf_1}}, *apply for*). Note that, it is difficult to statistically determine the implicit conflicts since such conflicts can only be detected for the coming access request during the system running.

It has been shown that any two rules can be compared by using the method of attribute complementation [17]. The distinction between the *e*xplicit and implicit conflicting rules just lies in the different representations , while the nature of both is the same. According to whether or not the conflicts happen, all the rules in the policy can be categorized into two classifications as follows:

(1) Probable-conflicting rules ar *and* ar'*need* meet all of the following conditions, where op*()*, d*()*, att*()*, and ap*()*, respectively, represent the corresponding functions or actions with respect to their prefixes:

$$\begin{aligned}
&(a) op(ar) \cap op(ar') \neq \varnothing,\\
&(b) d(ar) \neq d(ar'),\\
&(c)\ \begin{array}{l}\forall \text{att} \in (\text{att}(S,ar), att(O,ar)), \exists \text{att}' \in (att(S,ar'), \text{att}(O,ar')):\\ \text{att} = \text{att}',\end{array}\\
&(d) \forall ap \in (ap(S,ar), ap(O,ar)), \exists ap' \in (ap(S,ar'), ap(O,ar')):\ (\text{att}(ap) = \text{att}(ap')), (ap \cap ap' \neq \varnothing).
\end{aligned} \tag{7}$$

(2) Nonconflicting rules *ar* and *ar*' need meet one of the following conditions:

$$\begin{aligned}
&(a).\ d(ar) = d(ar'),\\
&(b).\ op(ar) \cap op(ar') = \varnothing,\\
&(c). \exists ap \in (ap(S,ar)\ or\ \ ap(O,ar)), \exists ap' \in (ap(S,ar')or\ \ ap(O,ar')):\\
&(att(ap) = att(ap'))\ and\ (ap \cap ap' = \varnothing).
\end{aligned} \tag{8}$$

3.3. Novel Method of Conflict Detections. Different rules may satisfy the same access request because of the flexibility and powerful expressiveness of the ABAC mechanism. The rules with both the *permitted* and *denied* values of the access decision, however, can cause contradictory access results. Therefore, it is necessary to study how to detect and resolve the existing conflict problems.

To comprehensively detect the conflict problems, we first propose the definitions of conflict-relation satisfiability and conflict probability as follows.

Definition 1. Conflict-relation satisfiability

The satisfiability of conflict relations is a function *satisfied*(*ar*, *conf*), which is used to check whether or not the

Table 3: Description of ABAC rules.

Rule	Attribute condition of the subject	Attribute condition of the object	Operation	Decision
ar_1	*role* = {*cashier, junior*}	*general business* = {*deposit, withdrawal*}	*perform, withdraw*	*permitted*
ar_2	*role* = {*general employee, manager*}	*loan business* = {*house*}	*apply for*	*denied*
ar_3	*role* = {*general employee*} and *work_year* < 20	*benefit business* = {bf_1, bf_3}	*apply for*	*denied*
ar_4	*trustness_level* > 12	*card business* = {$card_2$}	*apply for,withdraw*	*permitted*
ar_5	*trustness_level* > 8 and *work_year* > 10	*benefit business* = {bf_1, bf_2}	*apply for,withdraw*	*permitted*
ar_6	*trustness_level* < 10 and *work_year* < 15	*benefit business* = {bf_1, bf_2}	*apply for*	*denied*

rule *ar* could satisfy the specific conflict relation *conf*, where $ar \in P$, $conf \in Conf\_Set$. $Conf\_Set = Single\_Conf\_Set \cup Cross\_Conf\_Set$, which contains all the conflict relations between a single attribute and multiple attributes using the ABCL specification method. It can be formalized as

$$\forall ar \in P, \exists \text{conf} \in \text{Cons\_Set}:$$
$$\text{satisfied}(ar, \text{conf}) = \begin{cases} \text{true, if conf is satisfied} \\ \text{false, if conf is not satisfied} \end{cases}. \quad (9)$$

Definition 2. Conflict probability

Given any two conflicting rules ar_i and ar_j and suppose that there are n common attributes $att_1, att_2, \ldots, att_n$ existing in both the rules. If the access request is uncertain, then the conflict probability between ar_i and ar_j is denoted as

$$\text{Prob}\left(ar_i, ar_j\right) = P^C\left(\text{att}_1\right) \times P^C\left(\text{att}_2\right) \times \cdots \times P^C\left(\text{att}_n\right) = \prod_{att} P^C(\text{att}), \quad (10)$$

where $P^C(\text{att}) = \text{sim}(ap_i, ap_j) = |ap_i \cap ap_j| / |ap_i \cup ap_j|$; ap_i and ap_j are the attribute expressions of the common attribute *att* in ar_i and ar_j, respectively; the Jaccard coefficient $sim(ap_i, ap_j)$ of statistics, which aims to identify sample clusters, is used to measure the similarity between ap_i and ap_j.

According to the classification representation for the conflicting rules, as well as Definitions 1 and 2, we take the initial policy rules constructed by the conventional policy-engineering method as input and present the process of the conflict detections in Algorithm 1.

In the algorithm, we first create and initialize a temporary policy set P' and a result set *CP* of *conflicting-rule pairs* in lines 1 and 2. Next, for each rule *ar* in P', we check whether or not *ar* could satisfy a specific conflict relation *conf* (lines 3–7). If the *satisfied* function returns *false*, which indicates that some constraint is not yet satisfied, then rule *ar* is removed from the candidate policy set. Then, based on the descriptions of the probable-conflicting rules and non-conflicting ones, (lines 8–21) examine and calculate the conflict probability of each rule pair (ar_i and ar_j), in order to detect the *conflicting-rule pairs*, which provides quantitative evaluation criteria for resolving the conflicts and formulating the policies.

4. Experimental Analysis

Experiments are carried out in order to evaluate the performance of the satisfiability of the conflict relations as well as the conflict detections. All the experiments are compiled and run under the Java environment.

4.1. Performance Evaluations for the Satisfiability of Conflict Relations. We employ the real-world and public-available datasets from research [1], in order to construct initial ABAC policies using the Xu-Stoller [7] or VisMAP [9] method, while studying the formulation of the conflict relations, such as the SOD constraints and cardinality constraints. We utilize the Rel-SAT model counter [24] to generate constraint policies from the given SOD constraints.

To simulate the actual scenarios while meeting the security requirements, we implement the experiments in the initial policy-engineering system and adopt the same experimental setup with research [1], including the number of users and the scale of the policy. Figure 2 presents the performances of the proposal using different policy sets.

Two conclusions can be observed in Figure 2. First, as the number of users increases, the execution time does not vary obviously and tends to grow linearly for each given policy. Second, if the number of users remains unchanged, the execution time varies obviously as the scale of the policy changes. Specifically, when the scale of the policy is set at 20, the execution time is always close to 0.04 s, which remains almost stable. However, if the number of users remains constant and is set at 40, the time varies from 0.02 s to 0.04 s. This is because the larger the scale of the policy is, the more verification time for the SOD constraints generated by the Rel-SAT tool will be.

To further demonstrate the efficiency of the conflict-relation satisfiability, we present the following evaluation measure:

Reconstructed ratio (*RR*): It is used to quantitatively evaluate the satisfiability for the constraints during the formulation of conflict-relation sets, which can be denoted as: $RR = |SSAV'|/|SSAV|$, where the $SSAV'$ represents the relationship of the reconstructed attribute assignments that can satisfy the conflict relations, and the $SSAV$ represents the relationship of the initial attribute assignments.

Then, we take the cardinality constraint and SOD constraints as inputs, repeatedly implement the experiments on the real-world datasets, such as University and Healthcare, and output the median values of the experimental results as shown in Figure 3.

Figure 3 shows that the reconstructed ratio *RR* of attributes varies as the threshold of the cardinality constraint varies, where the scale number of the *Conf_set* respectively takes 100, 200, 300, and 400, and the number of attributes in the $SSAV$ constraint set is fixed. It can be observed that the

Input: the set *Conf_Set* of conflict relations and the initial policy set $P = \{ar_1, ar_2, \ldots, ar_n\}$, where $ar_i = (S_i^{ap}, O_i^{ap}, op_i, d_i)$.
Output: the result *set* CP *of conflicting-rule pairs, such as* (ar_i and ar_j) *and the conflict probability* Prob(ar_i and ar_j).

```
(1) Initialize CP = ∅;
(2) Create and initialize a temporary policy set P' = P;
(3) for each ar in P' do
(4)   if ∃conf ∈ Conf_Set: satisfied(ar, conf) = = false then
(5)     P' = P'\{ar};
(6)   end if
(7) end for
(8) for each rule pair (ar_i, ar_j) in P' do
(9)   if (d(ar_i)! = d(ar_j))∧(op(ar_i)∩op(ar_j)! = ∅) then
(10)    for each ap in S_i^ap or O_i^ap of ar_i do
(11)      for each ap' in S_j^ap or O_j^ap of ar_j do
(12)        if (att(ap) = = att(ap'))∧(ap∩ap' = = ∅) then
(13)          continue;
(14)        else
(15)          CP=CP∪{(ar_i, ar_j)};
(16)          calculate Prob(ar_i, ar_j);
(17)        end if
(18)      end for
(19)    end for
(20) end if
(21) end for
```

Algorithm 1: Novel method of conflict detection.

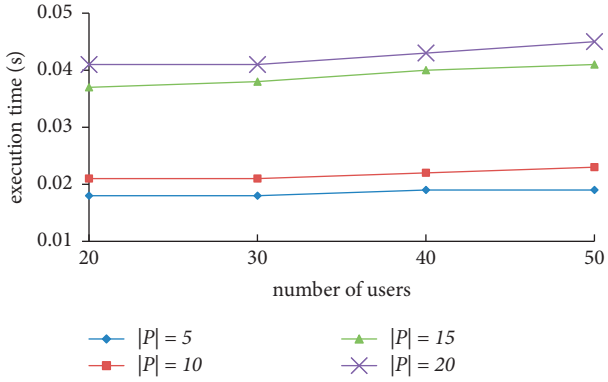


Figure 2: Execution time using different policies.

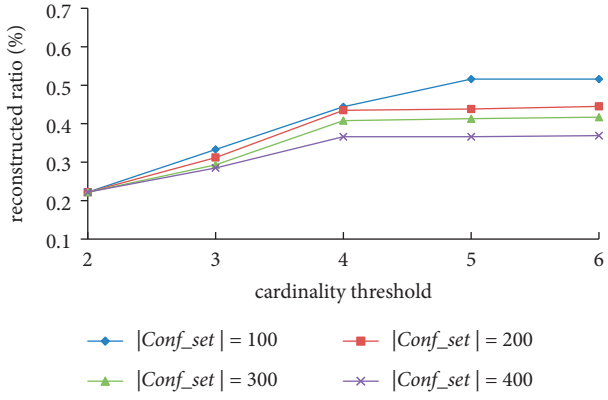


Figure 3: Evaluations of satisfiability using different constraints.

reconstructed ratio first increases significantly and then varies slightly as the constraint threshold increases. This is because the reconstructed ratio is positively correlative to the cardinality constraint, while the saturation will be present when the threshold of the cardinality constraint reaches a certain value. It can be also observed that the reconstructed ratio decreases with the increasing number of cardinality constraints, which is because of the restriction of the constraints on the attribute assignments.

4.2. Performance Evaluations for the Conflict Detections. Next, we use the Shu method [16] and Liu method [17], in order to obtain the synthetic datasets. Specifically, we assume that all the attributes belong to the natural-number type and the value ranges change from 1 to 100 since different attributes can take different values. The number of attribute conditions in each access request follows a normal distribution, and the total number of attributes is less than 40. The value of any attribute-expression condition varies between an upper bound and a low bound and follows a uniform distribution. According to the actual functional requirements of the business organization, we first design a generator to automatically generate various access requests, as well as, 30 different policy sets that are separated into 5 groups for the usage scenario, as shown in Table 4.

To demonstrate the effectiveness of the proposal in the phase of conflict detection, we propose two evaluation metrics as follows:

(1) Average conflicting number $Avg\_N(CP)$: it can be denoted as $Avg\_N(CP) = 1/M \sum_{i=1}^{M} N(ar_i)$, where

Table 4: Descriptions of the generated ABAC rule set.

Group	Policy set	Maximal length of the rule ($\|RL\|$)	Policy scale
1	P_1–P_6	3	100, 200, 400, 700, 1000, 1500
2	P_7–P_{12}	7	100, 200, 400, 700, 1000, 1500
3	P_{13}–P_{18}	11	100, 200, 400, 700, 1000, 1500
4	P_{19}–P_{24}	15	100, 200, 400, 700, 1000, 1500
5	P_{25}–P_{30}	20	100, 200, 400, 700, 1000, 1500

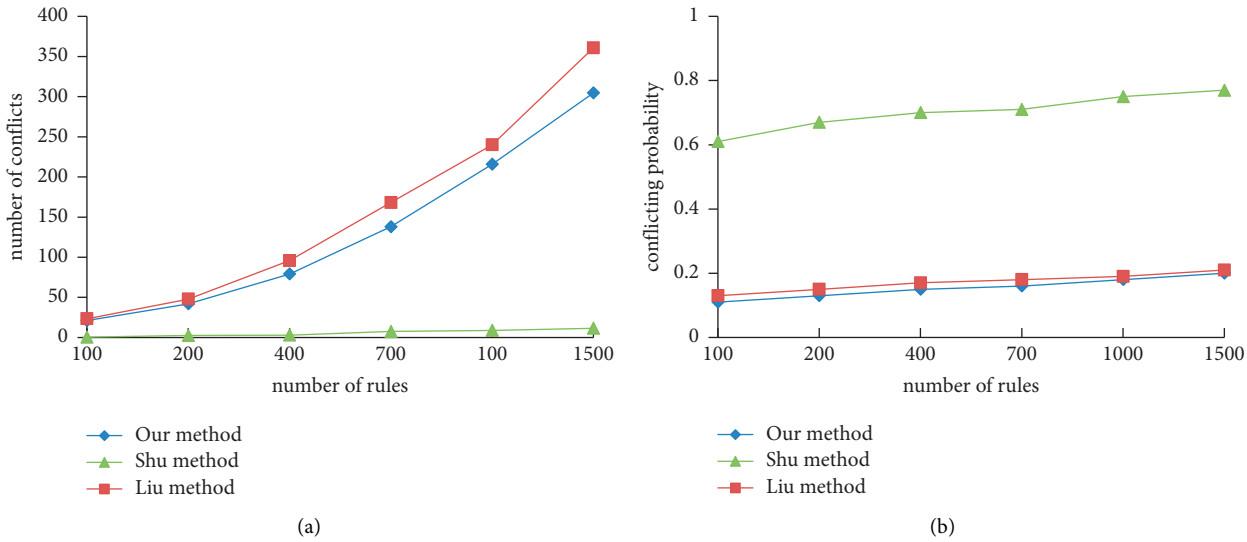


Figure 4: Performance comparisons when $|RL| = 3$. (a) Conflict number. (b) Conflict probability.

$N(ar_i)$ represents the conflicting number of rule ar_i, and M represents the scale number of set CP

(2) Average conflicting probability $Avg\_P(CP)$: it can be denoted as $Avg\_P(CP) = 1/M \sum_{i=1}^{M} (1 - \prod_{(ar_i, ar_j) \in CP} (1 - Prob(ar_i, ar_j))$, where $\text{Prob}(ar_i, ar_j)$ represents the conflict probability between ar_i and ar_j, and M represents the scale number of set CP

We implement experiments on the generated policy sets, calculate $Avg\_N(CP)$ and $Avg\_P(CP)$, and compare the performance of our method with the results of the Liu method and Shu method as shown in Figures 4–8.

Figure 4(a) shows that the average conflict number for P_1–P_6 policies vary with the increasing number of rules when the maximal length of the rule is set at 3. Specifically, the average conflicting number using our method varies from 21.17 to 304.63, which increases remarkably as the number of rules increases from 100 to 1500. Similarly, the result of the Liu method also increases significantly from 23.29 to 360.78 as the number of rules varies. However, the result of the Shu method tends to grow linearly and varies gradually from 0.45 to 11.5, which is a very small proportion of the actual conflicting rules. Obviously, the conflicting number using the Shu method is far less than those of the other two methods, which is not applicable to the actual requirements of organizations. This is because the Shu method only considers the attribute expressions and rules with the same identifiers. The rules with different attribute identifiers, however, are not taken into account using such a method. In fact, most conflicts occur among the implicit conflicting rules, which tend to be ignored and are not easy to be detected. To resolve this issue, both the Liu method and our method perform better and can detect the implicit conflicting results, and the conflicting number using our method is less than that of the Liu method. This is because the probable-conflicting rules violating the attribute-conflict relations are first removed using our method, before actually detecting conflicts. Furthermore, for the values of $|RL|$ taking 7, 11, 15, and 20, the varying tendencies of the average conflicting rules are presented in other figures, which are similar to that of figure 4(a). Notice that the results of both the Liu method and our method decrease as the number of the attribute-expression conditions increases. Thus, it is suggested to choose as many attribute conditions as possible, in order to formulate more flexible policies while reducing the number of probable conflicts.

Figure 4(b) shows the average conflict probability for the first 6 policies in *Group* 1. It is observed that using our method and the Liu method, the conflicting probability is always lower than 0.2 as the number of rules varies, while the result of the Shu method exceeds 0.6. With an increase in the length of the attribute-expression conditions, the average conflict probability decreases remarkably as shown in the figures. For instance, its value remains around 10^{-2}, 10^{-3}, 10^{-5}, and 10^{-7} when $|RL|$ takes 7, 11, 15, and 20, respectively. Both our method and the Liu method perform better than

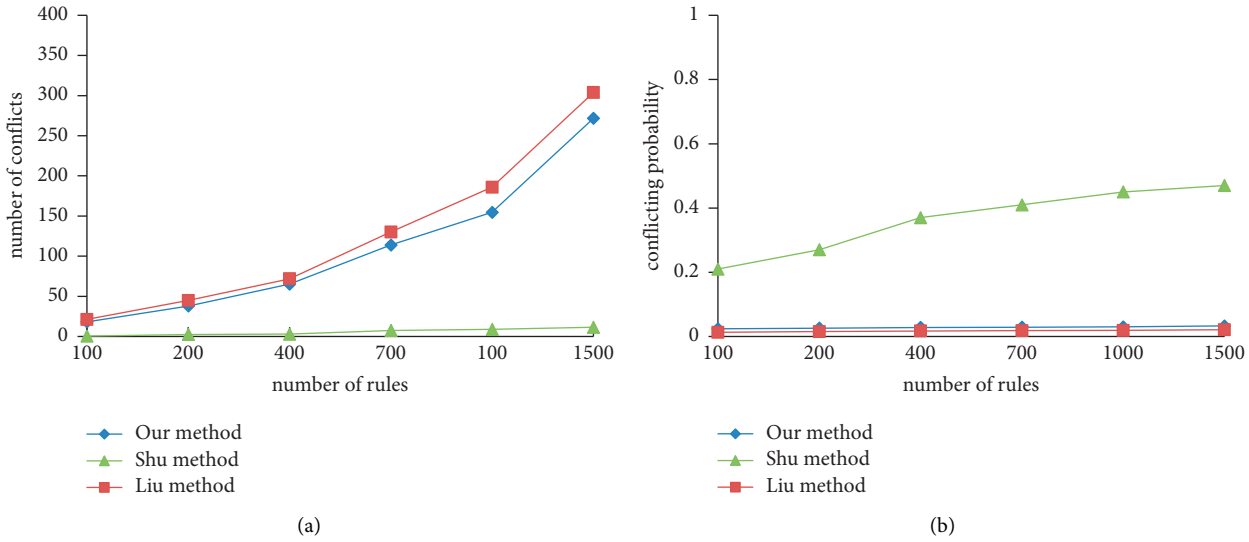


Figure 5: Performance comparisons when $|RL| = 7$. (a) Conflict number. (b) Conflict probability.

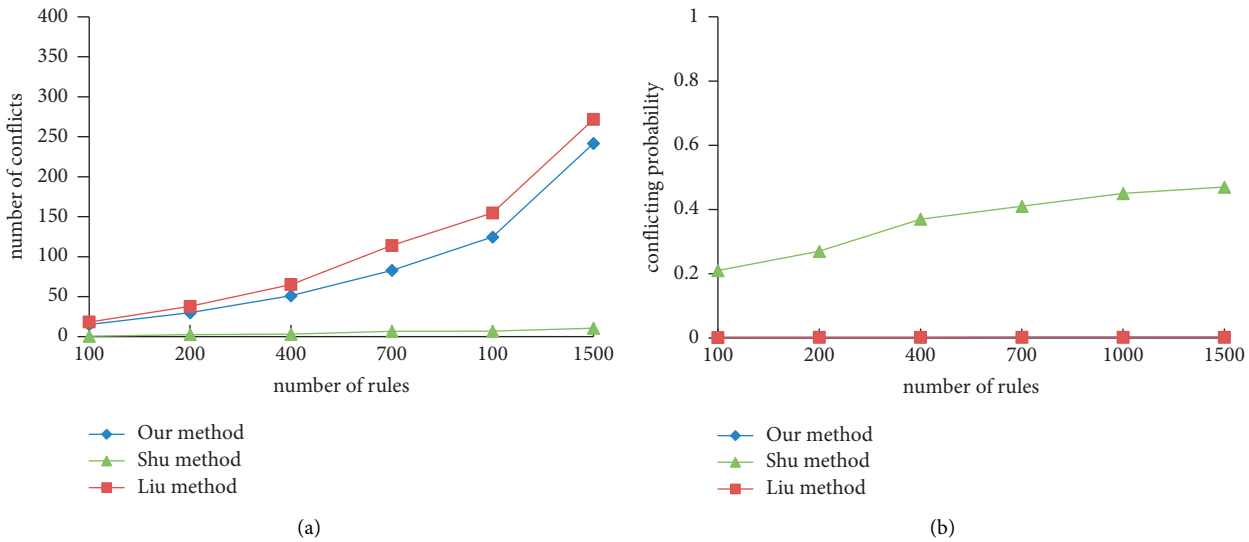


Figure 6: Performance comparisons when $|RL| = 11$. (a) Conflict number. (b) Conflict probability.

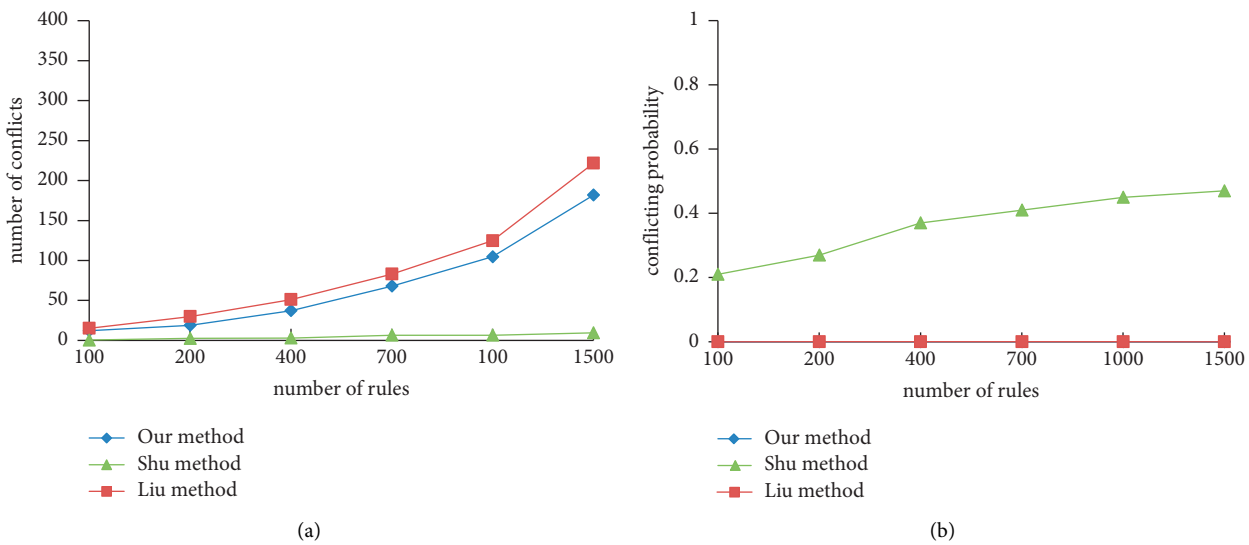


Figure 7: Performance comparisons when $|RL| = 15$. (a) Conflict number. (b) Conflict probability.

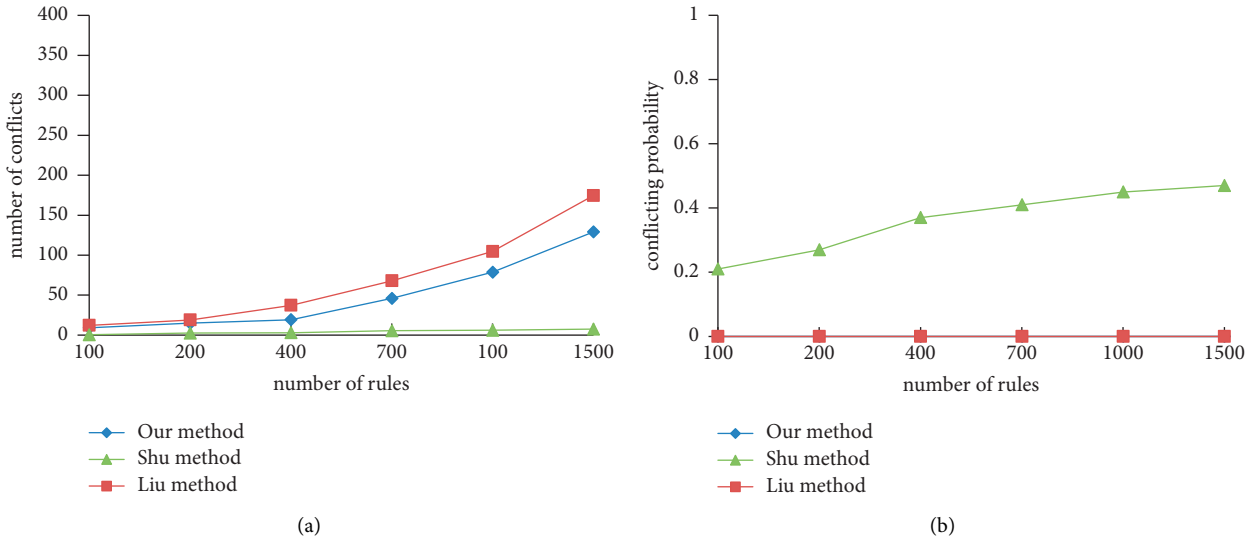


FIGURE 8: Performance comparisons when $|RL| = 20$. (a) Conflict number. (b) Conflict probability.

the Shu method from the viewpoint of conflicting probability. Furthermore, it is observed that the varying length of the attribute-expression conditions has a greater effect than that of the policy scale. Thus, it is also advised to use as many attribute conditions as possible, in order to improve the policy quality while reducing the conflicting probability.

5. Conclusions

A novel policy evaluation method, called ABPE_CSL&CD, was proposed in this study. According to the requirement descriptions of usage scenarios, we first utilized the attribute-based constraints specification language to formulate and specify the conflict relations among attributes, proposed the satisfiability of conflict relations, and categorized the conflicting rules into two classifications. Then, we presented the evaluation criteria on conflicting rules and attribute-conflict relations and proposed a novel algorithm for detecting conflicts. As a result, the proposed method flexibly suited the organizational requirements and comprehensively detected the ABAC conflict problems. The experiments on the real and synthetic datasets demonstrated that they could address the stated problems of improving the policy quality while reducing the conflicting number and conflicting probability. Our future work will focus on studying how to implement the ABPE_CSL&CD in other system scenarios such as the IoT, blockchain, and wireless sensor networks.

Data Availability

All the underlying data used to support the results of the study are included within the article.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the Key Scientific Research Project of Henan Province University.

References

[1] W. Sun, H. Su, and H. Xie, "Policy-engineering optimization with visual representation and separation-of-duty constraints in attribute-based access control," *Future Internet*, vol. 12, no. 10, p. 164, 2020.

[2] W. Sun, "Hybrid role-engineering optimization with multiple cardinality constraints using natural language processing and integer linear programming techniques," *Mobile Information Systems*, vol. 23, p. 1, 2022.

[3] S. Chakraborty, R. Sandhu, and R. Krishnan, "On the Feasibility of Attribute-Based Access Control Policy Mining," in *Proceedings of the 20th IEEE International Conference On Information Reuse And Integration For Data Science*, pp. 245–252, Los Angeles, CA, USA, August 2019.

[4] D. Servos and S. L. Osborn, "Current research and open problems in attribute-based access control," *ACM Computing Surveys*, vol. 49, no. 4, pp. 1–45, 2017.

[5] M. Narouei, H. Khanpour, H. Takabi, N. Parde, and R. D. Nielsen, "Towards a top-down policy engineering framework for attribute-based access control," in *proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies*, pp. 103–114, Indianapolis, IN, USA, June 2017.

[6] D. Mocanu, F. Turkmen, and A. Liotta, "Towards ABAC Policy Mining from Logs with Deep Learning," in *Proceedings of the 18th International Multiconference*, pp. 124–128, Ljubljana, Slovenia, October, 2015.

[7] Z. Xu and S. D. Stoller, "Mining attribute-based access control policies," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 5, pp. 533–545, 2015.

[8] S. Das, S. Sural, J. Vaidya, and V. Atluri, "Poster: using gini impurity to mine attribute-based access control policies with environment attributes," in *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies*, pp. 213–215, Indianapolis, IN, USA, 2018 June.

[9] S. Das, S. Sural, J. Vaidya, V. Atluri, and G. Rigoll, "VisMAP: Visual Mining of Attribute-Based Access Control Policies," in *Proceedings of the 15th International Conference on Information Systems Security*, pp. 16–20, Hyderabad, India, December, 2019.

[10] E. C. Lupu and M. Sloman, "Conflicts in policy-based distributed systems management," *IEEE Transactions on Software Engineering*, vol. 25, no. 6, pp. 852–869, 1999.

[11] J. C. Royer and A. S. De Oliveira, "AAL and static conflict detection in policy," *International Conference on Cryptology and Network Security*, pp. 367–382, Springer, Heidelberg, Germany, 2016.

[12] A. A. Jabal, M. Davari, E. Bertino et al., "Methods and tools for policy analysis," *ACM Computing Surveys*, vol. 51, no. 6, pp. 1–35, 2019.

[13] M. St-Martin and A. P. Felty, "A Verified Algorithm for Detecting Conflicts in XACML Access Control Rules," in *proceedings of the 5th ACM SIGPLAN Conference On Certified Programs And Proofs*, pp. 166–175, ACM, St. Petersburg FL USA, January 2016.

[14] M. Rezvani, D. Rajaratnam, A. Ignjatovic, M. Pagnucco, and S. Jha, "Analyzing XACML policies using answer set programming," *International Journal of Information Security*, vol. 18, no. 4, pp. 465–479, 2019.

[15] G. Zheng and Y. Xiao, "A Research on Conflicts Detection in ABAC Policy," in *7th International Conference On Computer Science And Network Technology*, pp. 408–412, IEEE, Dalian, China, October 2019.

[16] C. chun Shu, E. Y. Yang, and A. E. Arenas, "Detecting conflicts in ABAC policies with rule reduction and binary-search techniques," in *Proceedings of the 2009 IEEE International Symposium on Policies for Distributed Systems and Networks*, pp. 182–185, IEEE, London, UK, July 2009.

[17] G. Liu, W. Pei, Y. Tian, C. Liu, and S. Li, "A novel conflict detection method for ABAC security policies," *Journal of Industrial Information Integration*, vol. 22, Article ID 100200, 2021.

[18] A. Roy, S. Sural, A. K. Majumdar, J. Vaidya, and V. Atluri, "Enabling workforce optimization in constrained attribute based access control systems," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 4, pp. 1901–1913, 2021.

[19] X. Jin, R. Krishnan, and R. Sandhu, "A Unified Attribute-Based Access Control Model Covering DAC, MAC and RBAC," in *26th Annual IFIP WG 11.3 Conference On Data And Applications Security And Privacy XXVI*, pp. 41–55, Paris, France, July 2012.

[20] K. Z. Bijon, R. Krishnan, and R. Sandhu, "Towards an attribute based constraints specification language," in *Proceedings of the 2013 International Conference on Social Computing*, pp. 108–113, Washington, DC, USA, September, 2013.

[21] N. Helil and K. Rahman, "Attribute based access control constraint based on subject similarity," in *Proceedings of the 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications*, pp. 226–229, IEEE, Ottawa, ON, Canada, September 2014.

[22] S. Jha, S. Sural, V. Atluri, and J. Vaidya, "Specification and verification of separation of duty constraints in attribute-based access control," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 897–911, 2018.

[23] M. Alohaly, H. Takabi, and E. Blanco, "Towards an Automated Extraction of ABAC Constraints from Natural Language Policies," in *Proceedings of the 34th IFIP TC 11 International Conference On ICT Systems Security And Privacy Protection*, pp. 105–119, Lisbon, Portugal, June, 2019.

[24] J. B. Roberto and S. C. Robert, "Using CSP Look-Back Techniques to Solve Real-World SAT Instances," in *Proceedings of the 14th National Conference On Artificial Intelligence And Ninth Innovative Applications Of Artificial Intelligence Conference*, pp. 27–31, Providence, Rhode Island, USA, July, 1997.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 7267012, 10 pages
https://doi.org/10.1155/2022/7267012

Research Article

AI-Based Music Recommendation Algorithm under Heterogeneous Network Platform

Huan Wang

Jilin Normal University, Siping 136000, China

Correspondence should be addressed to Huan Wang; wanghuan@jlnu.edu.cn

Received 10 June 2022; Revised 20 July 2022; Accepted 26 July 2022; Published 2 September 2022

Academic Editor: Xingsi Xue

Music service is one of the diversified network services offered by people in the Internet era. Various music websites provide many tracks to meet people's music needs. Hundreds of millions of music of various genres at home and abroad, and there is a severe problem of information asymmetry between users and music. As a branch of the information filtering system, the recommendation system can predict users' preferences, increase flow, and drive consumption. A personalized music recommendation system can effectively provide people with a list of favorite tracks. Recently, many researchers have paid attention to heterogeneous networks because of their rich semantics information. Research has confirmed that rich relationship information in heterogeneous networks can improve the recommendation effect. Therefore, under the platform of a heterogeneous network, this paper divides the digraph set of track characteristics into several clusters with maximum heterogeneity, which makes the digraph of track characteristics in each cluster isomorphic to the maximum extent. When matching similarity, only searching in the cluster with the highest similarity to the target user can match a sufficient amount of applicable tracks, thus improving the efficiency of music recommendations to users. Experimental results show that the proposed algorithm has a high recall, precision, and F1 and can recommend personalized track lists to users to meet their music needs.

1. Introduction

Recently, the rapid advancements of mobile network technology have resulted in quick advancements of digital multimedia technology. Young people, especially students, have emerged as the primary consumers, and digital music has emerged as one of their preferred forms of consumer material [1, 2]. When users want specific music, they can easily search for it by entering information like title or artist, but when they do not have a clear query, that is, when they want the music system to give them music that meets their preferences without a clear goal, personalized music recommendation can be a better solution [3, 4].

The massive and huge music data generated in the music library undoubtedly exceeds the basic needs and bearing capacity of users, which leads to user information fatigue. In the face of the massive music data of the music library, ordinary music users often cannot quickly find the tracks that meet their preferences, and many personalized requirements for the music library recommended by others cannot be met [5–7]. Users cannot grasp or master a significant quantity of product information, or users have no specific aim in a certain sector but simply a broad desire, which is now an important problem to be handled [8]. The purpose of personalized music recommendation is to help users quickly screen out the music they are interested in from the vast music library. At present, most large-scale music portal websites have vast music libraries with a wide range of genres and styles of music, with new music being uploaded at a rapid rate every month. To begin with, the music library has hundreds of millions of tracks. Users will never have enough time to listen to all of the tunes before selecting their favorite. Second, music services are nonimmersive, and users can complete other things while listening to music. Music is only used as a background sound, which leads to vague demands of users, such as "recommend one or several nice tracks to me." The future market of music recommendation is very broad,

which fully meets the needs of users and can be accepted by users [9].

With the rise of music service, music recommendation technology related to music service also has a lot of research achievements [10–13]. Many music stations now offer not only basic music services but also the ability to push personalized playlists to users, notably Pandora and Last.fm. However, due to the uniqueness and sensibility of music itself, the contemporary suggestion results are lacking in personalized features and have a low coverage rate. People are more attracted to utilize mobile terminals for amusement and communication, thanks to the rapid growth of mobile terminal communication. A social network-based recommendation system has a clear business potential [14]. People are generally ready to share things with their friends on social networks, which include a large quantity of user information. Taking advantage of this link can boost the success rate of recommendations. At present, some mature social music platforms in China mainly use the data generated by users when they use the platform for social behaviors to calculate the similarities between users, thus predicting their interests and hobbies. On the social music platform, users can express their opinions on track messages and comments, from which we can extract the social tags that users place on track. The contents of these tags may include artists, music styles, music genres, users' current situations, feelings, moods, backgrounds, etc. These tags provide much information about the attributes of the track, as well as information about the scene when the user listens to the track, the user's immediate mood, ongoing activities or geographical location, etc., which are all helpful in our judgment.

Among many proposed recommendation algorithms, collaborative filtering algorithm is widely used, which uses the user's historical rating to recommend items that may be of interest to the user. However, due to the large number of items, users are often only able to rate a small number of items, resulting in data sparsity problems [15]. In addition, for a new user, due to the lack of rating information, it is difficult to make appropriate recommendations, so the recommendation system often faces the problem of cold start. To solve the problem of data sparsity and cold start, researchers have proposed many different algorithms. They have found that they can improve recommendations by exploiting relationships between users or items [16]. Since people with similar interests tend to like the same items, items with similar characteristics are more likely to be liked by the same users, while heterogeneous network contains rich relationship information, which can be used to improve the recommendation effect [17–19].

The main contributions of this paper are summarized as follows:

(1) Tag sequence-related attributes are mapped to a heterogeneous network.

(2) The digraph sets of track features are divided into several maximal isomorphic clusters so that each cluster's digraph of track feature is maximal isomorphic. In contrast, the digraph of track features in the different clusters differs. When matching similarity, sufficient applicable tracks can be matched only by querying in the cluster with the highest similarity with the target user, thus improving the efficiency of track recommendation for users.

The rest of this paper is organized as follows. In Section 2, we review the related works. The directed tag-based collaborative filtering algorithm in heterogeneous network is presented in Section 3. Experimental results are presented in Section 4. Section 5 concludes this paper.

2. Related Works

People's lives have created a demand for recommendation systems, and people want to consult other people's ideas while making judgments because of their highly socialized character. How to provide consumers with accurate and useful suggestions can help solve the problem of information overload while also benefiting the industry. Researchers have proposed a content-based recommendation algorithm and a collaborative filtering recommendation algorithm. These two recommendation algorithms, as well as their several modified variations, are now the most popular and widely utilized. Content-based recommendation is based on the user's historical behavior record to find the same as the item or has a certain context to recommend, requiring content information or expert annotation. The algorithm based on collaborative filtering has social characteristics and mainly recommends music matching users' interests and hobbies according to their interests, behavior records, and collection history. In [20], a new content-based recommendation method based on Gauss mixture model was proposed to improve the accuracy and sensitivity of probabilistic recommendation problems. In [21], a content-based recommendation algorithm based on convolution neural networks was proposed. To solve the cold start problem, in [22], the authors presented a rating forecasting framework, allowing the system to predict user ratings for unscripted music pieces, resulting in good recommendations. Currently, few recommendation systems consider users' interests and preferences at the same time. Considering each user's interaction, in [23], the authors proposed a user model and captured the user's interest. The traditional collaborative filtering recommendation algorithm has high computational complexity in calculating user similarity, leading to low recommendation efficiency. Therefore, in [24], the authors introduced the quantum computing theory to prepare the user score vector into a quantum state and calculate the similarity score in parallel. In [25], a hybrid web service recommendation method combining collaborative filtering and text content based on deep learning was proposed (HWSR-DL). In [26], a novel algorithm combining collaborative filtering and support vector machine was proposed to classify goods with positive feedback and negative feedback (CF-SVM). In [27], the authors proposed a new collaborative filtering method, which introduced information entropy and double clustering into collaborative filtering and extracted local dense rating module to deal with

the problems of data sparsity and low computational efficiency of traditional recommendation algorithms (IE-DC-CF).

In heterogeneous network platform, there are different types of links between nodes, which represent different kinds of relations and contain rich semantic information. How to calculate the similarity between nodes is an important problem in the process of extracting the relation information of heterogeneous networks [28–30]. With the rapid development of artificial intelligence machine learning in recent years, many new technologies have emerged. Researchers use various algorithms' characteristics to improve recommendation system performance. One is to guarantee the quality of recommendation results by data preprocessing. Some studies from quality evaluation and other aspects believe that the future recommendation system will become more perfect and mature.

3. Directed Tag-Based Collaborative Filtering Algorithm in Heterogeneous Network

Music tags might provide information about the track's information. Tags are primarily classified in tag-based music recommendation by the information relevance between tags [31]. However, because tags are separate from one another and disseminated in a distinct manner, we cannot know what users are thinking when they tag or classify music, and we cannot know their cognitive order of tags directly. To address this issue, we can make the tag directed to improve the situation. We may vectorize the time and times of users' activity data in music tagging to express the link between users, music, tags, and cognitive order and increase music recommendation accuracy.

Music stations offer services that allow users to comment on and rate music, thus keywords in user evaluations may be turned into music tags, and users can also choose from a list of optional tags for music tagging. The first few tags of a track are frequently named based on the artist's description, topic, and emotion, as well as the album's genre. When users play music, according to their perceptions of the music, they choose corresponding tags or, through the music, create their own tags to complete tagging. Users may have completely different feelings after repeatedly listening to the same song, and there may be many tags with significant differences. Each tagging of users will be recorded, and repeated tagging and comments will increase the weight.

The music tags issued by the music station are usually consecutive, and the more sophisticated the tags are, the more they match the track's features. This paper's data comes from the Million Song Dataset (MSD), which is an integration platform of music resources. It collected the data of seven well-known authoritative foreign music communities, sorted out and analyzed the data, and provided researchers with offline datasets and analysis results obtained by various algorithms. The offline dataset given by Last.fm [32] is mostly used for the optimization method data in this subsection. The offline dataset given by Last.fm is separated into a training set and a test set, with the training set accounting for 80% of the dataset and the test set accounting for 20%. This website is useful for comparison and discussion of subsequent studies since it gives tags and commonalities of track level. Figure 1 depicts the information for a specific piece of music on Last.fm.

There will be albums with various themes and playlists with various categories for artists. Each piece of music can be tagged by many people in the case of music. These tags may be similar or dissimilar, resulting in the music appearing in various playlists based on the tags. We may obtain the tag sequence for an artist's album as well as the tag sequence for music that has been tagged by various users. It should be noted that, in the playlist, users' cognition can be reflected in the sequence of tags. The more music can highlight the theme of the playlist, the more its tag and position should be placed in the front of the playlist. In the Last.fm dataset, track tags of users and artists are recorded, so the data is extremely large, with more than 200000 titles. The top 20 tags and their popularity are shown as Table 1.

Every user will have behaviors when listening to the tracks, such as playing, playing next, liking, looping, downloading, forwarding, and commenting, and the process time of the above behaviors will be recorded, which will make users become closely connected with tags. Tag information represents users' opinions on music, through which users are more likely to be interested in music, and in case of vague queries, we can use the tag of music to determine if the music is what they want.

We associate users with tag sequences according to certain rules and embody the relationship in terms of equations. U_i is the current i th user number. Assuming there are a total of x tags, the tag sequence is represented by $t_1, t_2, \ldots, t_x$, the tag sequence of the associated user is represented by a_{Ti}, the x th tag of the i th user tag is represented by $t_{i,x}$, and the associated user forms the following record as shown in

$$a_{Ti} = U_i, t_{i,1}, \ldots, t_{i,x}, \tag{1}$$

gSTU is used to represent the tag sequence after the user is associated, so when we obtain the different track and tag sequences that the user collects and tags; we can get the set of m tag sequences of the user, as shown in

$$a\mathrm{STU}_i = \left\{a_{Ti}^1, a_{Ti}^2, \ldots, a_{Ti}^i\right\}. \tag{2}$$

We use equation (3) to identify the sorted tags in Table 1.

$$a_{\mathrm{Tra}_j} = \mathrm{Track}_j, t_{i,1}, \ldots, t_{i,x}. \tag{3}$$

The above equation represents the sequence in which a certain track j is noted in a tracklist, and the x th label marked by this track is recorded as $a\_\mathrm{Tra}_j^n$. In addition, a piece of track may have multiple tags at the same time, and the track may appear in different track lists. Therefore, this tag sequence is recorded, as shown in

$$a\mathrm{STT}_j = \left\{a_{\mathrm{Tra}_j^1}, a_{\mathrm{Tra}_j^2}, \ldots, a_{\mathrm{Tra}_j^n}\right\}, \tag{4}$$

where $a_{\mathrm{Tra}_j^i}$ represents the i th tagging that track j has appeared in n tracklist tagging sequences.

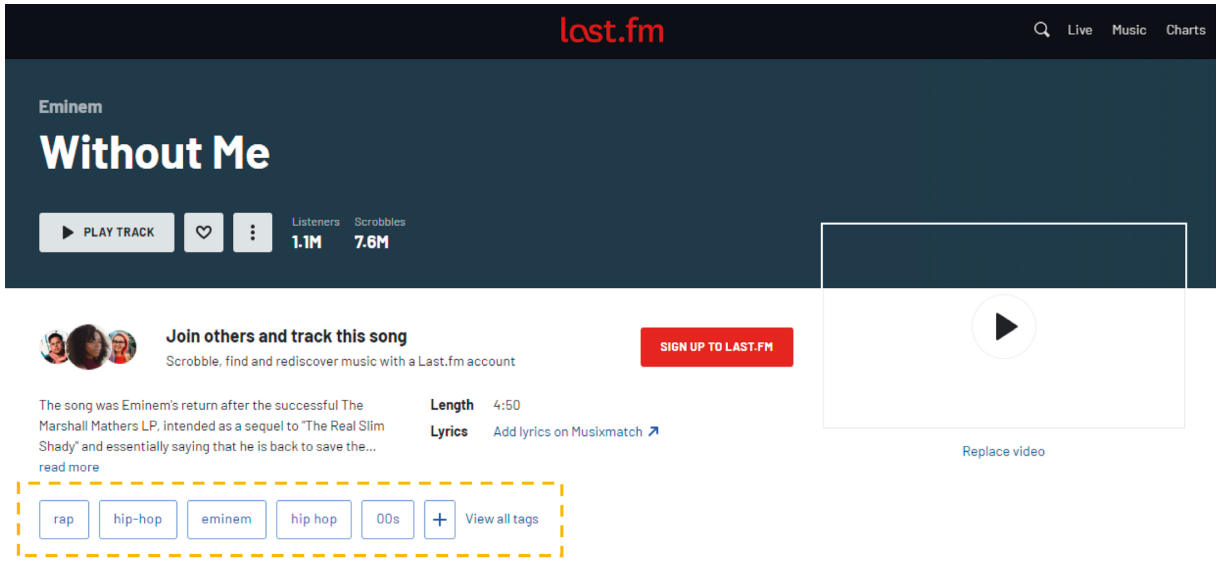


Figure 1: A track information from Last.fm.

Table 1: Parts tags and popularities.

Tag	Popularity
Hip-hop	101072
Rap	69159
Pop	55777
West coast	48175
Kendrick lamar	46720
Conscious hip hop	42564
Compton	39951
Jazz rap	31598
Trap	33618
West coast hip hop	31234
California	30433
Rnb	30125
Baby keem	29124
American	27810

For a heterogeneous network G, we assign a_{T_i}, $a\mathrm{STU}_i$, a_{Tra_k}, and $a\mathrm{STT}_j$ to G, that is, $G = (a_{T_i}, a\mathrm{STU}_i, a_{\mathrm{Tra}_k}, a_{\mathrm{Tra}_k})$. a_{T_i} corresponds to the vertex of G in the directed graph of heterogeneous information network, $a\mathrm{STU}_i$ corresponds to the edge of G, a_Tra_k corresponds to the vertex type of G, and $a\mathrm{STT}_j$ corresponds to the edge type of G.

The feature digraph of heterogeneous network is established by associating users with tags and track with tags [33]. Clustering was established for the digraph of track feature in heterogeneous network, and the track clustering was completed by using the clustering algorithm to obtain the central digraph of each cluster. The isomorphism degree of the cluster center digraph and user feature digraph in heterogeneous network is calculated, and then the isomorphism degree of the track feature digraph and user feature digraph in the cluster center digraph meeting the threshold value in heterogeneous network is calculated in turn to obtain the final result.

Through the above steps, we obtain the user's interest feature digraph, complete the clustering division of the track feature digraph, and obtain the feature digraph of its cluster center. The user's digraph of interest features is matched with the digraph of cluster center feature one by one, and the clusters whose isomorphism reaches the threshold are selected. Then the user's digraph of interest features is matched with the digraph of track feature in the cluster one by one, and the results are sorted.

When the total amount of track is small, the recommendation result of a single cluster may not be able to meet the needs of users. Therefore, when matching in this case, the critical value of isomorphism is taken as the definition, and the clusters with isomorphism higher than the critical value are stored in a new cluster. It is assumed that the critical value of isomorphism is γ, and the higher the critical value is, the higher the requirement for isomorphism is. Here, according to the average number of tags in the dataset, the isomorphism critical value is set at four to make track recommendation. When TopN is recommended, users will switch according to the scene and mood when playing tracks [34]. They will switch quickly if they are unsatisfied with the tracks when listening. If five or six tracks recommended in a row cannot satisfy users, they may even give up the recommendation and choose again. Considering that most users who listen to music using the recommendation list need a piece of background music or music that fits their mood in their spare time, the length len of the recommendation list is set to 25.

Our ultimate goal is to create a TopN list of music recommendation for U_i, that is, to calculate whether the isomorphism of the digraph of track feature and the digraph of user interest feature in the track cluster meets the recommendation requirements. If so, it is merged into a collection. Therefore, we need to create an adjacent clustering set $C_{\mathrm{neighbour}} = \{C_{n1}, C_{n2}, \ldots, C_{nN}\}$, where N represents the total number of clustering clusters. We store the set of the closest neighbors of the target digraph in this clustering set, then calculate the clustering center CC_i of the digraph of track features in heterogeneous networks, and calculate the

value of the isomorphism degree θ. Here, we compare the value of the isomorphism θ with the isomorphism threshold $\gamma = 4$ we set earlier. If $\theta > \gamma$, indicating that the isomorphism meets the recommended requirements, put the cluster of CC_i into $C_{\text{neighbour}}$; otherwise skip and judge the next cluster.

Put digraphs of track features in heterogeneous information network into recommendation list RecList[len] in descending order. Repeat this step until all the digraphs of track features in the cluster are judged and put into the recommendation list, and a complete recommendation list is obtained, in which clusters of the corresponding order are placed. By using the one-to-one relationship between digraph and track, we can get a track list created for users, which is defined as TrackRecList. Each digraph of track feature has its unique corresponding track, which is put into TrackRecList in order. We can get the final complete track recommendation list, which can be output as a result, so that users can get the final recommendation result. Due to the extensive data in this database, when analyzing the isomorphism degree of the digraph of user interest in the cluster center, we only need to query and match it with the clustering where the digraph with the highest similarity is located. Then we can make the music recommendation more efficiently.

The collaborative filtering algorithm based on user, track, tag, and tag sequence is the core of personalized music recommendation based on heterogeneous network designed in this paper, and its detailed process is shown in Figure 2.

4. Experimental Results and Performance Analysis

4.1. Dataset. There are 943347 matched tracks in the Last.fm dataset, with 505216 tracks having at least one tag, 584897 tracks having at least one comparable track, 522366 unique tags, and 8598630 track-tag pairs. We obtained 952067 tracks matching artists from the Last.fm dataset, stored them in the database for statistics, and then obtained the number of tags. Similarly, we counted the total number of tagged tracks, the number of users, the number of active users, the number of active tags, and the number of tracks with at least one tag. The specific statistics are shown in Table 2.

In Table 2, users who have tagged track for at least five times are defined as active users, and tags that have tagged track for at least five times are defined as active tags. Tagged track is track that has been tagged at least once. Based on the above data, the collaborative filtering algorithm model based on directed tags is used to carry out quantitative analysis of the model and construct a complete model. After our first step of screening, the tag noise has been reduced as far as possible. We have recorded the track tags and their corresponding occurrence frequency from the dataset to make recommendations.

We randomly selected 1000 users from the Last.fm dataset who were highly active users (with more than 10 tagging behaviors) and 1000 users from the normal active users (with 5–10 tagging times) as experimental objects. Users with low activity levels are not considered, because any recommendation system must be based on user data, and without user behavior data, it is impossible to provide users with accurate and satisfactory track recommendation services. Two groups of data extracted are used for the experiment, namely, High-Active User Dataset (HAUD) and Normal-Active User Dataset (NAUD). The specific data is shown in Table 3.

Using the playing time and tagging time recorded in the database, the first 80% users in the dataset were taken as the training set, and the rest were taken as the test set. Although the weight of each user is different, leading to differences in individual recommendation results, such an experiment is more appropriate to a real recommendation system on the whole, and the results are closer to real data.

In addition to the above algorithm reference, we set the length of the recommendation list to 5, 10, 15, 20, 25, 30, and 35, respectively, to consider the accuracy of the algorithm. The reason for this is that a short recommendations list is not persuasive, while a long recommendations list can lead to impatience, disgust, and poor results. We need to select the most appropriate length of track recommendation list through experiments. After setting up the control group, a fair standard is needed to evaluate the excellence of each algorithm. Generally, accuracy or recall is used for consideration, which is based on recommendation results, order, and actual item correlation value. Since the evaluation of these two metrics is not comprehensive, F1 score is introduced as a comprehensive evaluation standard for the experiment.

It is assumed that reco is the resource set with length n obtained from the recommended result, and real is the real resource set of the user. count_i is used to record whether the i th resource of reco is in real. If count_i is in the real, the value is 1; otherwise it is 0. To verify the performance of the proposed algorithm, three algorithms such as HWSR-DL [25], CF-SVM [26], and IE-DC-CF [27] are used as baselines.

4.2. Results and Analysis. In the experiment, considering that it is meaningless and has some side effects when recommending track lists to users, here we take {5, 10, 15, 20, 25, 30, 35} as seven values to test the list length of recommended tracks and then test the datasets HAUD and NAUD separately. Subsequently, DTCF-HN is used to represent directed tag-based collaborative filtering algorithm proposed in this paper. We compare the recall, accuracy, and F1 of each algorithm in two different test sets HAUD and NAUD with different n values.

As can be seen from Figure 3, with the increasing of recommendation list length, recall of all algorithms shows an increasing trend. The recall performance of all algorithms in HAUD dataset is superior to that in NAUD dataset, indicating that sufficient user behavior data information can make the algorithm perform better. Comparing the results of the two datasets, it can be seen that the recall of the DTCF-HN algorithm decreases a lot in the NAUD dataset, and the stability of the algorithm is slightly worse than that of the IE-DC-CF algorithm with a higher degree of data dependence, but it still has advantages over the other two algorithms. This

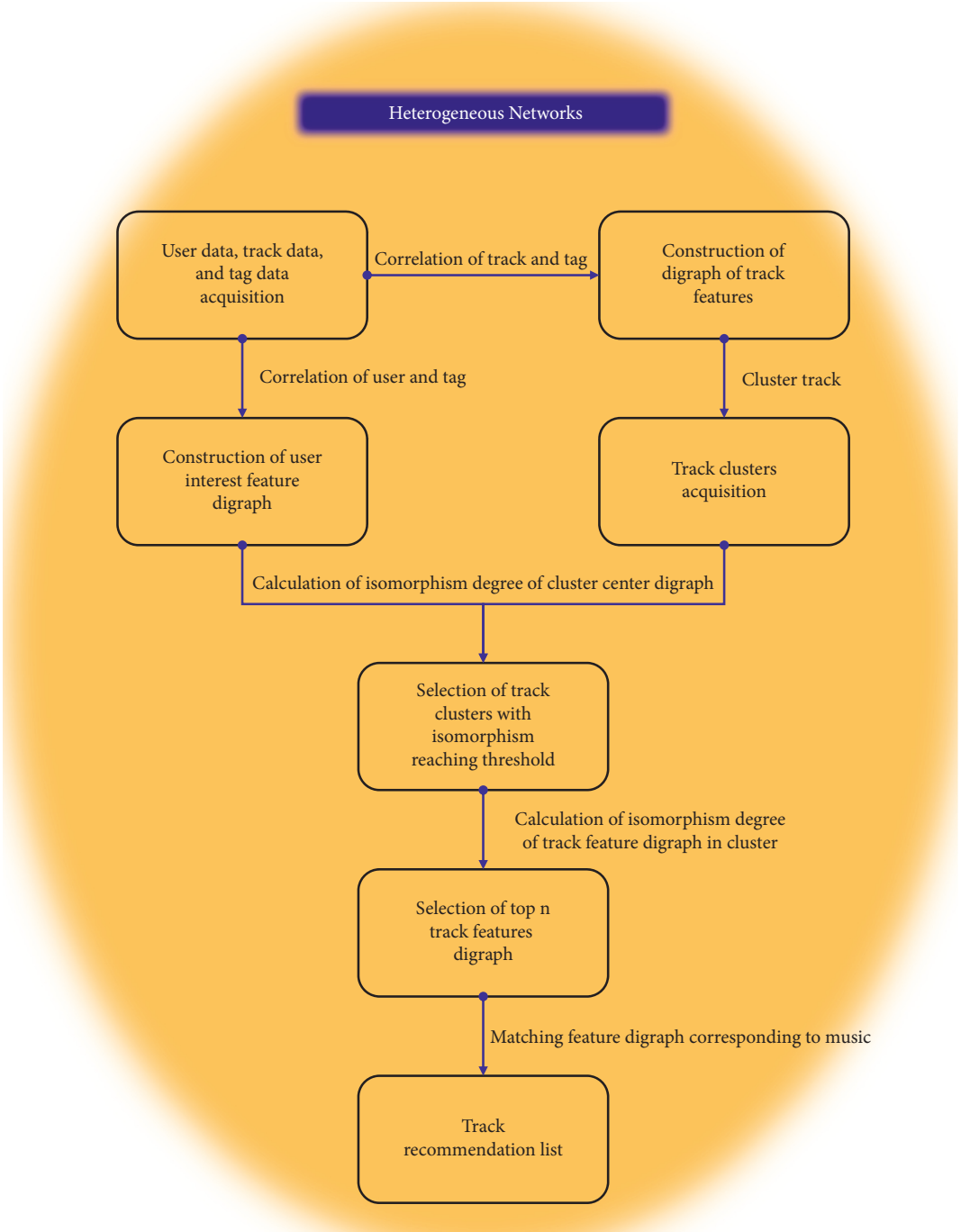


Figure 2: Music recommendation algorithm based on artificial intelligence under heterogeneous networks platform.

Table 2: Data statistics from Last.fm.

Number of users	Number of active users	Number of tags	Number of active tags	Number of tracks with tags
1573950	1157451	532648	128456	502635

Table 3: Information of HAUD and NAUD.

Dataset	Number of users	Number of tracks	Number of tags	Tagging times
HAUD	1000	20987	1982	15129
NAUD	1000	16534	1374	7651

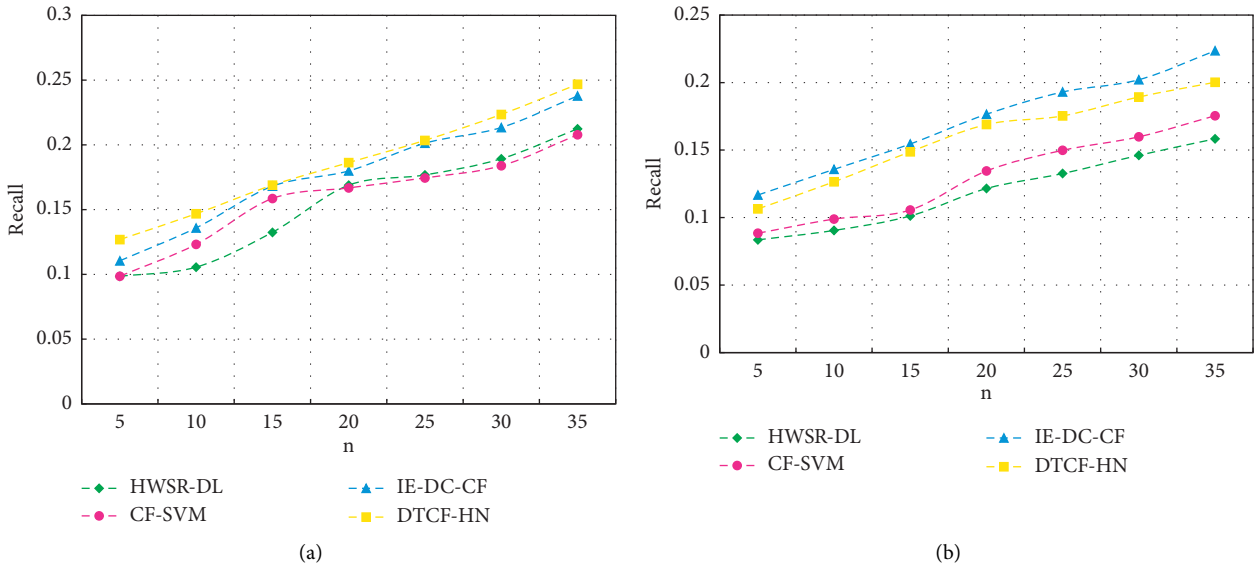


Figure 3: Recall of four algorithms in different datasets. (a) HAUD. (b) NAUD.

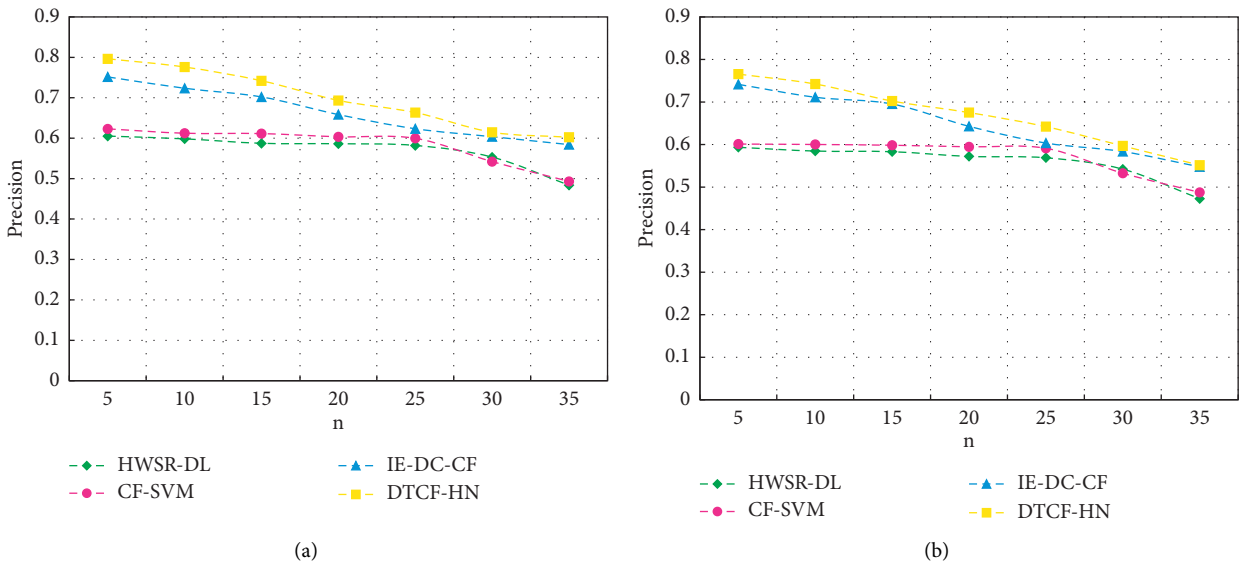


Figure 4: Precision of four algorithms in different datasets. (a) HAUD. (b) NAUD.

shows that DTCF-HN algorithm performs well and proves the hypothesis that tags have certain sequentiality.

As can be seen from Figure 4, with the increasing of the length of the recommendation list, the precision of the four algorithms shows a decreasing trend. HWSR-DL and CF-SVM algorithms show a relatively stable performance in the first recommendation list length of 5–25, and the precision begins to decline when n is greater than 25. The accuracy of DTCF-HN algorithm shows an accelerating trend in the process of decreasing, and the overall precision is good, indicating that the track with the sequence in the front can satisfy users more and proving that the track recommendation results provided by the algorithm have a relatively clear sequence. The performance of each algorithm on HAUD dataset is better than that on NAUD. DTCF-HN algorithm has the most obvious performance gap in the two datasets, indicating its stronger dependence on data. The main reason is that this algorithm focuses on mining the horizontal relationship between tag data, and the number of tagging behaviors has a significant impact on algorithm performance. In the NAUD dataset, the performance of DTCF-HN algorithm is slightly worse than that of IE-DC-CF algorithm, but it still has certain advantages compared with the other two algorithms, indicating that DTCF-HN algorithm performs well in accuracy.

As indicated in Figure 5, F1 of each algorithm shows a trend of increasing first and then decreasing. Compared with the datasets of HAUD and NAUD, the reduction of data volume leads to an increase in the recommended list length n required to reach the peak value of F1, which means that the

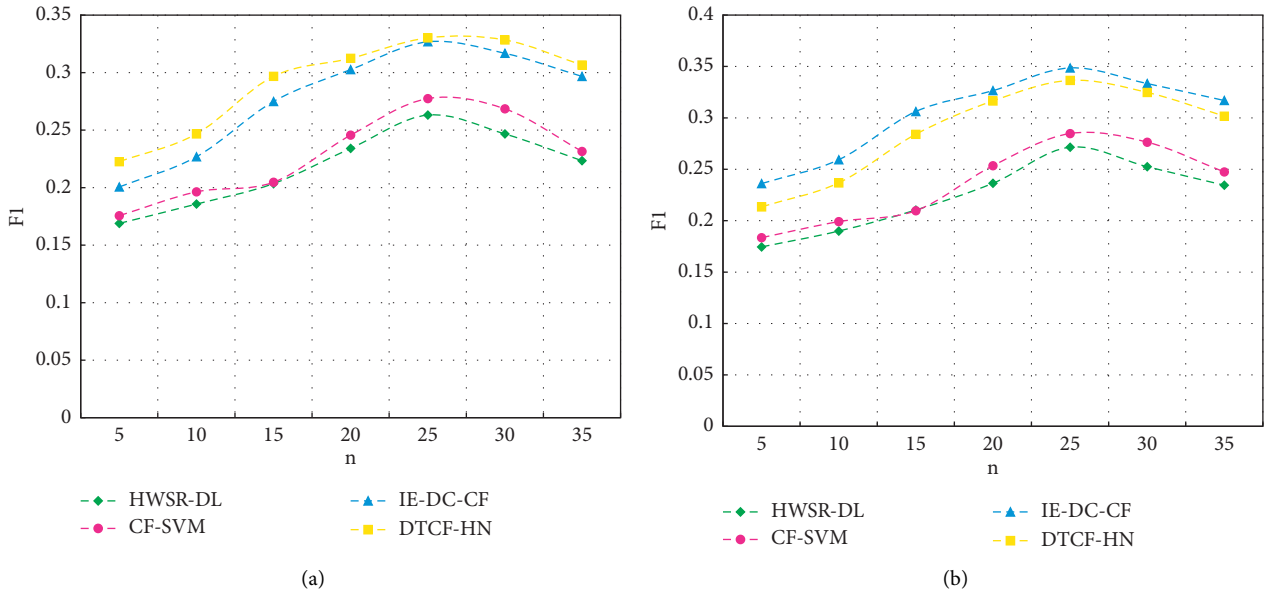


Figure 5: F1 of four algorithms in different datasets. (a) HAUD. (b) NAUD.

accuracy will decrease and the corresponding performance will decrease. In the experimental results, the performance of DTCF-HN and IE-DC-CF algorithm is significantly better than that of the other two algorithms. In the HAUD dataset, the performance of DTCF-HN is better than that of IE-DC-CF algorithm, but it is the opposite in NAUD. This is because DTCF-HN considers the sequential relationship between tag data and is more dependent on the amount of data. According to the above experimental results, except for the IE-DC-CF algorithm, compared with the other two algorithms, DTCF-HN has apparent advantages in performance and can provide users with more satisfactory recommendation results, which verifies the hypothesis that there is a specific sequence between tags and also indicates that DTCF-HN algorithm has good performance.

5. Conclusion

The rapid development of mobile terminals has made digital music mainstream, and major Internet companies have also increased their investment in the music field. Huge demand brings enormous traffic, so how to provide users with their favorite music in the massive music database has become the focus of competition among significant Internet music businesses. Therefore, music recommendation algorithm based on personalization has been developed for decades. There are countless outstanding researchers to provide music recommendation services by combining advanced mathematical statistics ideas with computers with high-speed processing power. Mainstream recommendation algorithms have advantages and disadvantages, and combining them will improve the recommendation effect. In the case that open datasets are relatively easy to obtain, the cost of a collaborative filtering algorithm is lower than that of the content-based algorithm, and it has a better effect on cluster recommendation. In this paper, the music feature digraph is clustered and divided, so there is an apparent distinction between the clusters. At the same time, it ensures that each cluster's music is isomorphic with the cluster center feature digraph to the greatest extent. When recommending track lists, it only needs to match the user feature digraph with the track in the cluster with the highest fitness. Experimental results show that the proposed algorithm has a high recall, precision, and F1 and can recommend personalized track lists to users to meet their music needs.

Through the analysis of the experimental results, it is proved that the algorithm has good performance, but there are still some shortcomings. We can further improve the performance through the following aspects.

(1) The algorithm proposed in this paper relies on user data and track data. If there is a problem with data sparsity, the recommendation result cannot meet the expectation. In the future, the content-based music recommendation can be integrated into the content-based music recommendation according to the semantics of lyrics, that is, using the hybrid model.

(2) The amount of data has a significant impact on the performance of the algorithm. Music recommendations that meet the requirements cannot be provided for users with few annotation behaviors. In future research, we can conduct in-depth longitudinal research on tags, such as mining emotional indicators and implicit semantic information in tags, to further improve the performance and stability of the algorithm.

(3) Construct more advanced computing frameworks, such as Spark, a distributed memory framework, and MapReduce, a framework in Hadoop, to increase the batch processing capacity of files. For tasks such as Last.fm, which have a large amount of data and require intensive computing, the use of distributed

frameworks can reduce time costs and improve iteration efficiency.

Data Availability

The data used to support the findings of this study are available upon the reasonable request.

Conflicts of Interest

The author declares that he has no conflicts of interest in this article.

References

[1] D. Arditi, "Music everywhere: setting a digital music trap," *Critical Sociology*, vol. 45, no. 4-5, pp. 617–630, 2019.

[2] D. Nakano, "Digital music, online outlets and their business models," *Brazilian Journal of Operations & Production Management*, vol. 16, no. 4, pp. 581–591, 2019.

[3] S. C. Brown and A. E. Krause, "Freedom of choice: examining music listening as a function of favorite music format," *Psychomusicology: Music, Mind, and Brain*, vol. 30, no. 2, pp. 88–102, 2020.

[4] S. Li, Q. Luo, L. Qiu, and S. Bandyopadhyay, "Optimal pricing model of digital music: subscription, ownership or mixed?" *Production and Operations Management*, vol. 29, no. 3, pp. 688–704, 2020.

[5] H. Im, H. Song, and J. Jung, "A survival analysis of songs on digital music platform," *Telematics and Informatics*, vol. 35, no. 6, pp. 1675–1686, 2018.

[6] T. Hodgson, "Quantifying music: imagined metrics in digital startup culture," *Culture, Theory and Critique*, vol. 61, no. 4, pp. 424–439, 2020.

[7] Z. Li and H. Ma, "Development and design of music education resource library based on web," in *Proceedings of the 2017 2nd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2017)*, pp. 1036–1040, March 2017.

[8] X. Zhang, X. Ding, and L. Ma, "The influences of information overload and social overload on intention to switch in social media," *Behaviour & Information Technology*, vol. 41, no. 2, pp. 228–241, 2020.

[9] Z. Cheng, J. Shen, and T. Mei, "Just-for-Me: an adaptive personalization system for location-aware social music recommendation, SIGIR'14," in *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1267-1268, Glasgow, United Kingdo, April 2014.

[10] S. Volokhin and E. Agichtein, "Towards intent-aware contextual music recommendation: initial experiments," *ACM/SIGIR Proceedings*, vol. 2018, pp. 1045–1048, 2018.

[11] M. Wang, Y. Xiao, W. Zheng, X. Jiao, and C. Hsu, "Tag-based personalized music recommendation," in *Proceedings of the 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN 2018)*, pp. 201–208, Yichang, China, February2018.

[12] D. Wang, S. Deng, S. Liu, and G. Xu, "Improving music recommendation using distributed representation," in *Proceedings of the 25th International Conference on World Wide Web (WWW'16 COMPANION)*, pp. 125-126, Québec, Montréal, Canada, April 2016.

[13] G. Zhong, H. Wang, and W. Jiao, "MusicCNNs: a new benchmark on content-based music recommendation," *Neural Information Processing*, pp. 394–405, Springer, Berlin, Germany, 2018.

[14] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: recommendation models, techniques, and application fields," *Electronics*, vol. 11, no. 1, p. 141, 2022.

[15] Y. Hu, W. Shi, H. Li, and X. Hu, "Mitigating data sparsity using similarity reinforcement-enhanced collaborative filtering," *ACM Transactions on Internet Technology*, vol. 17, no. 3, pp. 1–20, 2017.

[16] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: a short survey," in *Proceedings of the 2013 ASE/IEEE International Conference on Social Computing (SOCIALCOM)*, pp. 882–885, Alexandria, VA, USA, September 2013.

[17] Z. Zhao, X. Zhang, H. Zhou, C. Li, M. Gong, and Y. Wang, "HetNERec: heterogeneous network embedding based recommendation," *Knowledge-Based Systems*, vol. 204, Article ID 106218, 2020.

[18] C. Zhang, Z. Tang, B. Yu, Y. Xie, and K. Pan, "Deep heterogeneous network embedding based on Siamese Neural Networks," *Neurocomputing*, vol. 388, pp. 1–11, 2020.

[19] Y. Xie, B. Yu, S. Lv, C. Zhang, G. Wang, and M. Gong, "A survey on heterogeneous network representation learning," *Pattern Recognition*, vol. 116, Article ID 107936, 2021.

[20] N. Van Dat, P. Van Toan, and T. M. Thanh, "Solving distribution problems in content-based recommendation system with Gaussian mixture model," *Applied Intelligence*, vol. 52, no. 2, pp. 1602–1614, 2021.

[21] J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, "A content-based recommendation algorithm for learning resources," *Multimedia Systems*, vol. 24, no. 2, pp. 163–173, 2018.

[22] K. Okada, M. Kanamaru, P. X. Tan, and E. Kamioka, "Exploiting MUSIC model to solve cold-start user problem in content-based music recommender systems," *Intelligent Decision Technologies*, vol. 15, no. 4, pp. 749–760, 2022.

[23] B. R. Cami, H. Hassanpour, and H. Mashayekhi, "User preferences modeling using dirichlet process mixture model for a content-based recommender system," *Knowledge-Based Systems*, vol. 163, pp. 644–655, 2019.

[24] X. Wang, R. Wang, D. Li, D. Adu-Gyamfi, and Y. Zhu, "QCF: quantum collaborative filtering recommendation algorithm," *International Journal of Theoretical Physics*, vol. 58, no. 7, pp. 2235–2243, 2019.

[25] R. Xiong, J. Wang, N. Zhang, and Y. Ma, "Deep hybrid collaborative filtering for Web service recommendation," *Expert Systems with Applications*, vol. 110, pp. 191–205, 2018.

[26] D. Chang, H. Y. Gui, R. Fan, Z. Z. Fan, and J. Tian, "Application of improved collaborative filtering in the recommendation of E-commerce commodities," *International Journal of Computers, Communications & Control*, vol. 14, no. 4, pp. 489–502, 2019.

[27] M. Jiang, Z. Zhang, J. Jiang, Q. Wang, and Z. Pei, "A collaborative filtering recommendation algorithm based on information theory and bi-clustering," *Neural Computing & Applications*, vol. 31, no. 12, pp. 8279–8287, 2019.

[28] H. Shakibian and N. M. Charkari, "Statistical similarity measures for link prediction in heterogeneous complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 501, pp. 248–263, 2018.

[29] W. Wen, D. D. Zeng, J. Bai, K. Zhao, and Z. Li, "Learning embeddings based on global structural similarity in heterogeneous networks," *IEEE Intelligent Systems*, vol. 36, no. 6, pp. 13–22, 2021.

[30] X. Gao, J. Chen, Z. Zhan, and S. Yang, "Learning heterogeneous information network embeddings via relational triplet network," *Neurocomputing*, vol. 412, pp. 31–41, 2020.
[31] E. Zheng, G. Y. Kondo, S. Zilora, and Q. Yu, "Tag-aware dynamic music recommendation," *Expert Systems with Applications*, vol. 106, pp. 244–251, 2018.
[32] Z. Bodo and E. Szilagyi, "Connecting the Last.fm dataset to LyricWiki and MusicBrainz. Lyrics-Based experiments in genre classification," *Acta Universitatis Sapientiae, Informatica*, vol. 10, no. 2, pp. 158–182, 2018.
[33] Y. Sun and H. Zhao, "Eigenvalue-based entropy and spectrum of bipartite digraph," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3451–3462, 2022.
[34] Y. Xiao, K. Zhang, M. Xu, and Y. Liu, "A social TopN recommendation scheme based on grey forecast model," *Journal of Grey System*, vol. 30, no. 4, pp. 78–96, 2018.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 3022280, 11 pages
https://doi.org/10.1155/2022/3022280

Research Article

Research on Optimization of Cross-Border e-Commerce Logistics Distribution Network in the Context of Artificial Intelligence

Jihua Shi

Business School, Guangzhou College of Technology and Business, Guangzhou 510850, China

Correspondence should be addressed to Jihua Shi; shijihua@gzgs.edu.cn

Received 28 June 2022; Revised 18 July 2022; Accepted 18 August 2022; Published 28 August 2022

Academic Editor: Chin-Ling Chen

The continuous innovation of artificial intelligence technology has led to industrial upgrading and industry transformation in various industries, and e-commerce logistics has borne the brunt. Artificial intelligence relies on the intelligence of a large number of operations and decisions in the process of logistics operations, as well as the integration of transport, storage, distribution, packaging, loading and unloading, and other aspects of the production process and the hierarchy of the system, which has become an important engine to promote the upgrading of logistics equipment and technology, the innovation of production links and processes, the change in the structure of supply and demand of traditional logistics jobs, and the renewal of the traditional form of logistics. The application of artificial intelligence and the development of a new generation of information technology, including big data, has opened the era of intelligent logistics. And the continuous deepening of China's foreign trade makes the demand for cross-border e-commerce logistics surge, and the importance of cross-border e-commerce logistics has been rapidly highlighted. Hence, it becomes urgent to construct a cross-line Internet business coordinated operations improvement model by consolidating man-made consciousness innovation. In view of investigating the activity techniques of cross-line online business coordinated factors, this paper advances the improvement methodology of cross-line web-based business operations advancement way with regard to man-made reasoning. By concentrating on the ongoing circumstance and issues of cross-line web-based business coordinated factor circulation module and consolidating the idea of wise strategies, the appropriation stage, the way transportation stage, and freight conveyance phase of Internet business coordinated factor dissemination activity are advanced, in order to accomplish the motivation behind lessening the expense of end dispersion, working on the effectiveness of dissemination, and expanding consumer loyalty.

1. Introduction

In recent years, e-commerce, with its own advantages, has received strong support from the government and high attention from commercial enterprises, which makes the survival environment of e-commerce have been greatly improved and the development rate is very amazing [1, 2]. The logistics industry, which is closely connected with e-commerce, has also been greatly affected. Compared with the traditional logistics, the "new logistics" in the e-commerce environment presents unprecedented new features: logistics management informationization, logistics management networking, logistics management automation, logistics management intelligence, and logistics management flexibility [3, 4]. Among them, the management of intelligence, through artificial intelligence technology, to reduce human physical and mental labor, such as logistics engineering operation research problems, is reflected in the intelligence of logistics [5]. These characteristics mark the development of e-commerce logistics toward a more intelligent and intelligent direction.

Cross-line web-based business operations, both cross-line online business and worldwide coordinated factor credits, not just have a critical supporting job for the improvement of cross-line Internet business and global strategies, but also are a significant piece of global exchange. Starting from 2020, with the developing improvement of China's unfamiliar exchange, the size of cross-line online business planned operations proceeds to grow, and the interest for cross-line web-based business coordinated

factors on the planet has shown a dangerous development. China is a significant individual from the global exchange family, and commodity exchange possesses a significant position in the development of the public economy [6, 7]. To upgrade the designation of cross-line online business operation assets and advance top to bottom participation among cross-line Internet business coordinated factors undertakings, China has laid out a cross-line web-based business strategy partnership in collaboration with globally prestigious cross-line web-based business planned operation ventures, meaning to give particular cross-line Internet business planned operation administrations for cross-line online business planned operation clients. The information shows that through the arrangement of cross-line online business strategy unions with numerous nations, China's cross-line web-based business planned operation industry working together has quickly worked on both scale and quality. Be that as it may, due to the not-really lengthy improvement time, China's cross-line web-based business coordinated factor industry is dispersed and wasteful. Subsequently, how to fabricate and upgrade the coordinated factors and conveyance network for cross-line Internet business improvement has turned into a practical issue that should be concentrated critically.

As we all know, the logistics mode is constantly innovated with the development of productivity, from "first-party logistics" to "third-party logistics," and information sharing and coordination gradually become the concept that cannot be ignored in the value chain bearing logistics mode [8, 9]. With the quick improvement of the Internet, the online business data trade stage infiltrates into this worth chain, planning the organic market creation and activity exercises, everything being equal, while storing network strategies, the board coordinates different assets in the chain through the center hub of web-based business stage. Among them, man-made consciousness improves the capability of strategies of the store network board and really upholds the activity of cross-line online business planned operation framework. With the solid backing of large information and cloud stage, the functional objective of chasing after the center hub of the online business stage as the center hub of the store network framework with the best general effectiveness is understood step by step. As of now, the qualities of web-based business strategies contain essentially informationization, organizing, and robotization, no matter what is firmly connected with man-made brainpower.

2. Analysis of the Current Situation of Cross-Border e-Commerce Logistics and Distribution Network

2.1. The Current Situation of Cross-Border e-Commerce Wisdom Logistics Application in China. Along with the development and application of these new generation information technologies, such as the Internet of Things, cloud computing, big data, remote sensing technology, and artificial intelligence, in the field of logistics in China, wisdom logistics has also developed rapidly, making China's logistics industry increasingly powerful [10, 11]. In the past few years, the total amount of social logistics is rising year by year. Robotic high-tech logistics equipment such as drones, unmanned warehouses, ground wolves, sky wolves, AGV trolleys, and robotic arms has been initially developed and applied, and new technologies such as cold chain technology, logistics sky eyes, smart cabinets, and AI robots have also driven the development of smart logistics. It can be seen that wisdom logistics in China will make more long-term development and progress in the future. Of course, for the time being, China's smart logistics is still in the initial stage, and there is great room for development.

2.2. Existing Problems. From the perspective of the three modules of the e-commerce logistics distribution system, there are some urgent problems in each module at present.

2.2.1. Distribution Aspect. On the whole, although China has made great development and progress in the application of infrastructure in logistics warehousing operations, it is still relatively backward compared with developed countries in logistics [12]. Among them, third-party logistics enterprises have 60%–70% of the warehouse or ordinary cottage warehouse, the application to automated warehousing enterprises is few, and the application to the characteristics of e-commerce logistics to achieve multivolume, small batch, short-cycle automated picking and out of the warehouse is even less. Among the warehouse handling tools, still with trolleys, ground cattle, manual handling vehicles, and basic equipment, modern storage equipment such as AGV trolleys and other handling tools is rare. In terms of software, most logistics enterprises purchase or independently developed information software, and the customer's information system is not compatible.

2.2.2. In-Transit Transportation. Transportation cost is still the highest part of the logistics distribution process. From trunk transportation to terminal distribution, various transportation methods are not smoothly connected, which leads to high transportation costs. The loss, leakage, and damage of goods during transportation occur frequently, so how to ensure the tracking of goods and real-time monitoring of transportation vehicles has become an important problem to be solved [13, 14]. At the same time, the construction of a large number of end distribution networks and the huge number of couriers, etc., will increase the cost of delivery. And the complex urban traffic conditions and the increase in the number of private cars lead to increased traffic congestion, which brings more uncertainty to logistics and distribution, resulting in low delivery efficiency of courier companies and reduced customer satisfaction.

2.2.3. Delivery of Goods. As the "last mile," there is a point to multipoint characteristics, and customer distribution is more dispersed, mainly concentrated in the district, schools, and office buildings, so there are often multiple courier delivery trolleys gathered, resulting in travel congestion,

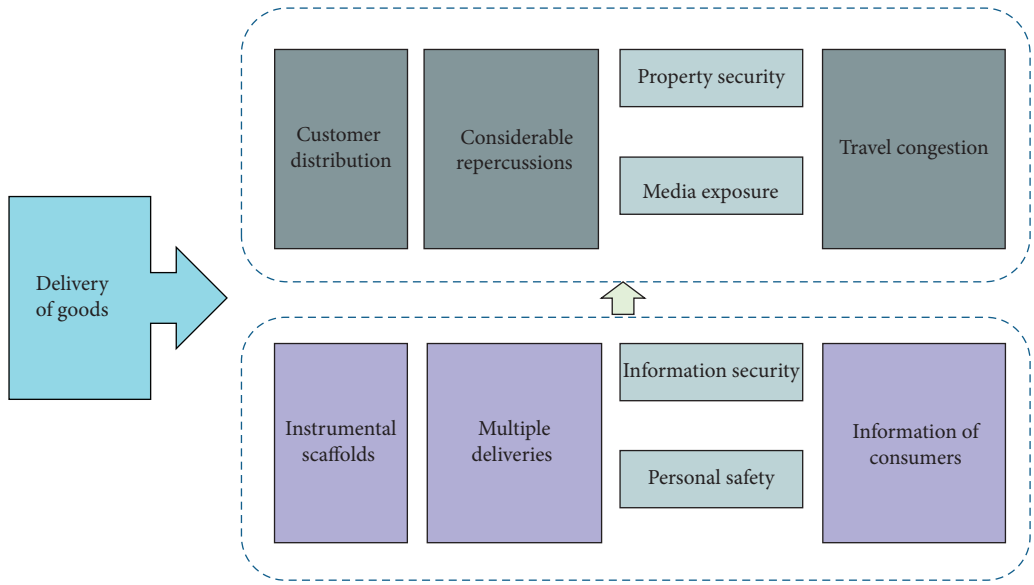


Figure 1: Technical flow chart of the delivery of goods.

bringing inconvenience to customers. At the same time, the phenomenon of delivery failure is more common, as customers work days in the unit and nonworking days at home; there may be a mismatch with the dispatcher delivery time; it is necessary to make multiple deliveries or received by others, which in turn may bring the problem of unclear responsibility; once the package is an anomaly, customer complaints will then increase. In addition, the recent media exposure of the leakage of personal information of consumers has also caused considerable repercussions; consumers are very worried about personal safety, property security, and information security. Figure 1 illustrates the technical flow chart of the delivery of goods.

3. Artificial Intelligence Technology

Artificial intelligence (AI) belongs to a branch of computer science, which tries to understand the nature of intelligence and produce a new intelligent machine that can respond precisely in a way similar to human intelligence [15]. Artificial intelligence can simulate individual consciousness and thinking information processes. Although AI is not human intelligence, it can think like a human and even surpass human intelligence, which shows that it is a very practical discipline. In addition, artificial intelligence is a developmental discipline with many uncertainties, which makes it rich in mystery and exploration.

3.1. Artificial Intelligence+ Logistics Integration Trend. The integration of artificial intelligence technology and the logistics industry can not only empower traditional logistics effectively but also create novel logistics service projects. Traditional logistics activities are refined into four major parts: inventory, warehousing, transportation, and distribution, while the new logistics service has practical service capabilities and effectiveness such as high synergy at the management level, profit maximization at the economic level, instantaneous in time, green in environmental protection, and intelligent and sensitive in individual experience. Regarding the empowerment of artificial intelligence, it mainly involves mechanical learning, the Internet of Things, unmanned systems, and other high-tech information technology fields.

3.2. Application Value of Artificial Intelligence in Logistics

3.2.1. Realization of Unmanned Delivery. Unmanned delivery vehicles are mainly used in express or just-in-time logistics distribution, using low-speed driving unmanned vehicles, the essence of which is basically no different from an autonomous driving system, which is composed of modules such as environment perception, vehicle positioning, path planning decision, vehicle control, and vehicle execution. The unmanned delivery vehicle receives and processes data through multisensor data fusion such as LIDAR, ultrasonic radar, cameras, and inertial sensors, then identifies and understands dynamic and static information such as roads, signs, pedestrians, vehicles, and the environment through machine learning and deep learning, and then makes route planning and behavioral decisions through differential positioning and high precision maps [16, 17]. In a nutshell, these cloud services provide data, high precision maps, algorithm updates, and background monitoring for the unmanned vehicles, and finally, the control system and execution system of the unmanned vehicles perform navigation, avoidance, acceleration, turning, braking, and other operations. The current application scenarios of unmanned delivery machines are still very limited. Unmanned delivery machines are more sensitive to objective conditions such as environment and climate and more complex route planning and algorithms, require more types of sensors, and avoid crowds and buildings for safety reasons. Due to policy restrictions, unmanned delivery aircraft are currently mostly used for delivery in remote or closed areas and emergency delivery.

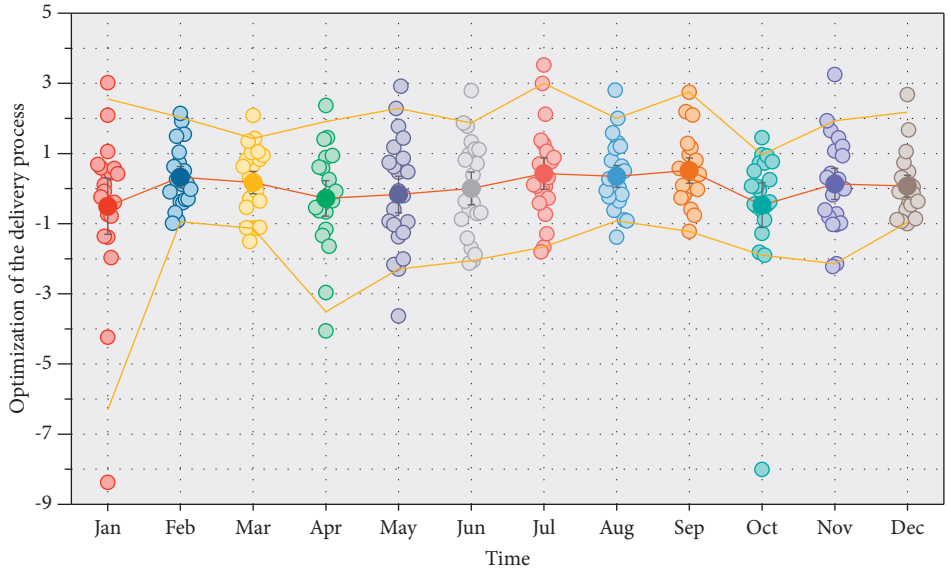


Figure 2: Dotted line diagram of optimized scheduling efficiency at different times of the year.

3.2.2. Optimize the Delivery Process. The "large information + calculation" in planned operations will compute the conveyance course to the most sensible and give information, high accuracy map, calculation update, and foundation checking for automated vehicles, so the products can arrive at the clients more securely and rapidly. "Large Data + Algorithm" can gather information on the rider's direction, constant climate, and conveyance business, consolidate with continuous information from the server farm to dissect through streamlining calculations and booking calculations, progressively plan the ideal way, cooperate with messengers continuously and proficiently, immediately report issues in conveyance, and lastly anticipate the dispatch's conveyance time through AI. Figure 2 shows the optimized scheduling efficiency through artificial intelligence prediction scheduling at different times of the year.

The large information stage will be coordinated with the venture's data framework to give a precise picture of the messenger and the freight proprietor and enter the planned operations data through PC acknowledgment innovation to dispatch and get merchandise, keeping away from pointless utilization of time and labor because of conceivable conveyance blunders brought about by manual request input [18]. Sending the pickup code to clients makes the precision of data conveyance higher and takes care of issues, for example, clients not having the option to track down the pickup warning in time, while likewise keeping clients educated regarding package pickups. Man-made brainpower will likewise channel and dissect the large information as indicated by the prerequisites of the errand and suggest the undertaking shrewdly for the deliverer as per the limit, model, area, and available energy of the deliverer.

Optimization of the delivery process (O) is chosen to measure the comparability between the conveyance course and pointless utilization of time and labor, and the formula is

$$T_i = \sum_i \ln\left[\left(-\frac{d_{pi}^2}{2S_p^2\sigma_i^2}\right)\delta\right],$$
$$O = \frac{T_i}{\sum \delta}, \tag{1}$$
$$O = \frac{\sum_i \ln\left[\left(-d_{pi}^2/2S_p^2\sigma_i^2\right)\delta\right]}{\sum \delta},$$

where p is the ID of the certain merchandise, i is the ID of the key point of conceivable conveyance blunders, d_{pi} denotes manual request input between the i-th pickup code by the p-th data conveyance, S_2 denotes available energy of the deliverer, and δ is the venture's data framework.

3.2.3. Intelligent Warehouse Management. In the 21st century, when human resources become more and more expensive, fully automated, high-efficiency intelligent management of the warehouse gradually emerged. Workers do not need to carry goods one by one but directly enter the goods number by the system to arrange the machine to store and find the goods, which is intelligent access to goods. The intelligent management mode allows users to participate in the management of goods and real-time monitoring of the status and location of goods through the network.

4. Artificial Intelligence Cross-Border e-Commerce Logistics and Distribution Optimization Ideas

The existing problems of low efficiency, poor quality, and waste of resources of cross-border e-commerce logistics distribution network can be solved by creating intelligent storage, intelligent transportation, and intelligent delivery

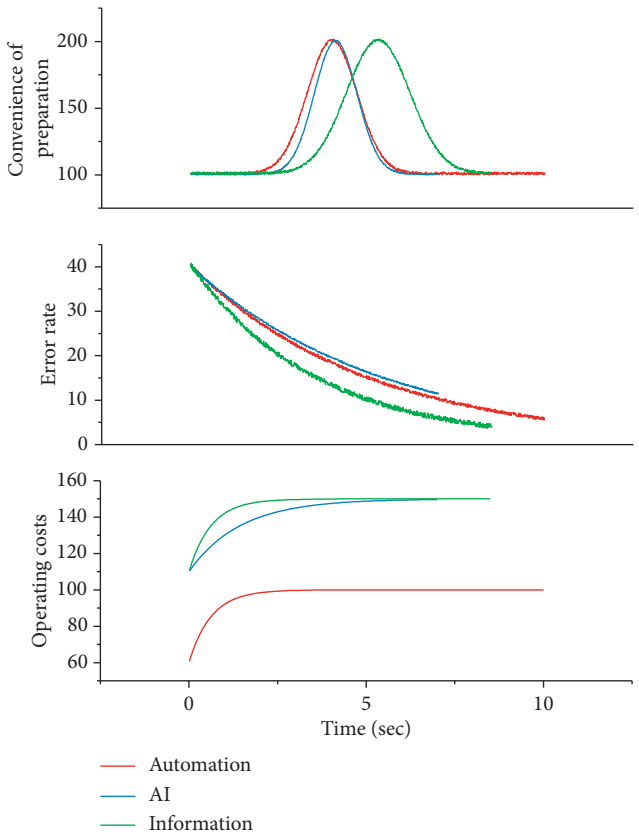


Figure 3: Curve diagram of artificial intelligence to improve the efficiency of distribution.

through high technology equipment, Internet of Things, cloud computing, big data, and other technologies.

4.1. Artificial Intelligence to Improve the Efficiency of Distribution. The operation in the warehouse includes receiving, shelves into the warehouse, storage in the warehouse storage, picking by order, out of the warehouse delivery, and several basic processes. The use of automation and AI equipment and information management software not only can improve efficiency but also reduce the error rate of manual operation, effectively improve the convenience of preparation of goods out of the warehouse, and reduce operating costs, as illustrated in Figure 3. The use of artificial intelligence technology will be helpful for warehousing as well as inventory management [19, 20]. At this stage, China's logistics delivery time requirements will be more stringent; once the logistics time of its goods is relatively long, then consumers are easy to produce dissatisfaction and other psychological emotions and take artificial intelligence technology and in-depth analysis of historical data, so as to better grasp the actual access law of inventory goods and dynamically to rectify the inventory so that it not only can better reduce the actual cost of warehousing costs but also effectively enhance the timeliness of warehouse work so that consumers are more satisfied.

(1) Receiving link can use natural navigation unmanned forklift. Unmanned forklifts can complete pallet handling operations, that is, the entire pallet of goods from the transport vehicle to the receiving area waiting for quality inspection into the warehouse. The natural navigation unmanned forklift does not need to install markers or reflectors and can perform self-positioning according to the information obtained from internal and external sensors during its movement, so as to achieve precise positioning and path planning of the unmanned forklift and complete the task of navigation. At the same time, intelligent depalletizing robots are used for depalletizing. This means that the goods placed on the transfer pallets are transported one box at a time to the conveyor belt. The use of depalletizing robots can greatly reduce the labor of workers, improve the speed of handling, and save labor costs. The latest intelligent robot can even realize the function of automatically adjusting the width of the arm for different sizes of boxes, without the need for warehouse personnel to adjust and calibrate the box type.

According to Wang et al. [21], unmanned forklift V_j is constructed to represent the entire pallet of goods from the transport vehicle to the receiving area Q to the information obtained from internal and external sensors (each information i corresponds to an intelligent depalletizing robot, representing u and v coordinates of the function of automatically adjusting the width of the arm for different sizes of boxes, respectively), and then the transfer pallets and warehouse personnel Z_k are fused to select labor of workers z_k from the predicted deviation h_k. The mathematical relationship is as follows.

$$V_j = \frac{\sum Q\left[u_j + Z_k\left(u_j\right) - u_i\right] h_k\left(u_j\right)}{\sum h_k\left(u_j\right)},$$

$$f_k\left(u_k\right) = \sum_j \frac{1}{\pi R^2} V_j, \tag{2}$$

$$f_k\left(x_k\right) = \sum_j \frac{1}{\pi R^2} \frac{\sum Q\left[u_j + Z_k\left(u_j\right) - u_i\right] h_k\left(u_j\right)}{\sum h_k\left(u_j\right)}.$$

(2) Unmanned handling vehicles mainly use electromagnetic or optical and other related principles to automate the deguidance device facility to carry out the project's marching work in a timely manner and transport the goods to the preset location points in accordance with its previously set related guidance path [22]. We take artificial intelligence algorithm, use intelligent equipment, timely cargo sweeping, transmission, and other various works, and improve the intelligent technology and automatic control technology so that the storage robot can go on its own operation, extensive use of neural networks and a series of algorithms, reasonable planning of good transport path, and accurate inference of its environmental change conditions, to achieve the purpose

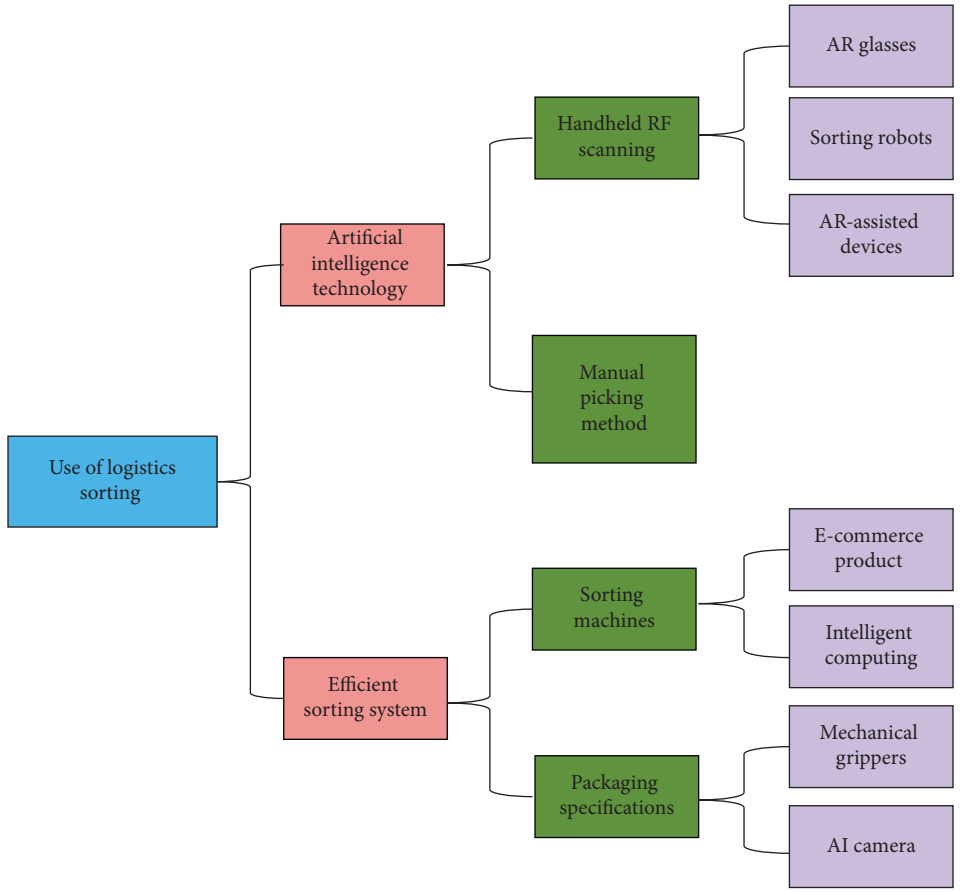


Figure 4: Organization frame composition of logistics sorting.

of unmanned. In the library, a storage link can be used in the common automated three-dimensional warehouse (AS/RS), the whole box of goods stored in the high level of the three-dimensional warehouse as a unit of pallet, through the aisle automatic stacker crane to complete the pallet access operations. And the part that needs to be unpacked and picked up at a later stage can be completed by an automated trolley.

(3) Use of Logistics Sorting

In the traditional form of manual sorting, the workload of the staff is too large for the staff to be able to cope with the problems. The adoption of artificial intelligence technology and the rational use of intelligent equipment and related technologies can build a more efficient sorting system, bringing new technologies such as conveyor sorting systems into practice and extending the actual scope of their use. Handheld RF scanning can be used in the goods picking process instead of the original manual picking method, and the handheld RF can accurately find the location of the goods according to the flashing lights of the goods level. Handheld RF scanning with AR-assisted technology is more intelligent and modern, and it is equipped with AR glasses to accurately and quickly understand the layout of the warehouse and the location of each storage position and can scan the goods through AR glasses, significantly improving operational efficiency. For secondary sorting, automatic sorters or sorting robots can be used. Automatic sorting machines are suitable for boxed or bagged goods with neat and uniform packaging specifications, and most domestic e-commerce goods are currently packaged in these two forms. The picking robot can automatically change mechanical grippers of different sizes through the camera and intelligent computing, which is suitable for the increasingly rich variety of e-commerce product packaging types (in Figure 4).

(4) Operating equipment with an automatic picking function can solve the function of order sorting and concentration according to delivery routes and stack the goods of the same customers centrally. The intelligent shipping sorting system can also pick and load according to the customer's delivery order, in the order of delivery routes from far to near, reducing the intermediate secondary handling process and improving the efficiency of outbound storage.

4.2. Artificial Intelligence to Improve Transport Efficiency. When the distribution logistics cannot better meet the sales demand of the network, the network goods cannot be transported to the hands of consumers in a timely manner,

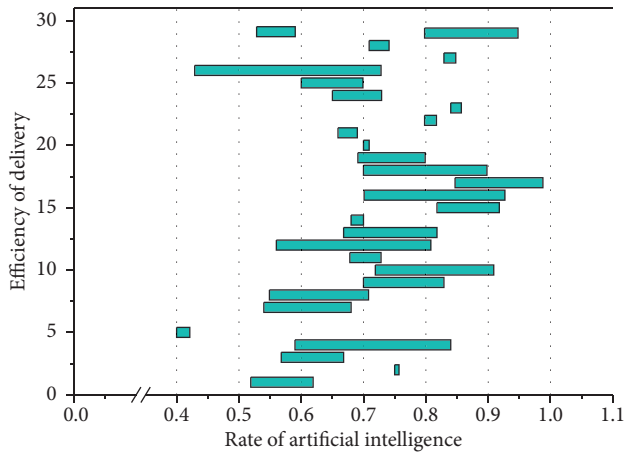


Figure 5: Horizontal column diagram of artificial intelligence to improve the efficiency of delivery.

and the arrival time node will also produce a delay, not to mention the completion of door-to-door commissioning as well as installation and other follow-up tasks, and the related service quality is poor [23, 24]. Therefore, when optimizing the logistics distribution network, we should build an integrated mode of sales and logistics distribution, integrate the offline logistics distribution data information with the online logistics sales goods information content, and optimize the information content to build a one-stop service chain to effectively break the traditional logistics and transportation mode. The shackles of the traditional logistics and transportation model are effectively broken, and the existence of blind spots is eliminated so that logistics and transportation can be developed in a more intelligent way.

Logistics management system using GIS technology and GPS positioning can realize GPS positioning service, human-computer interaction for logistics information management, real-time continuous positioning, monitoring, and data transmission of vehicles as well as the optimization of path selection and navigation of electronic maps, giving full play to the advantages of intelligent logistics [25]. The distribution center can use the vehicle path optimization system, with the best route, the shortest time, and the fastest speed to deliver goods to customers, greatly improving the efficiency of distribution, but also easing the city traffic congestion. Through the real-time tracking of vehicles, the vehicles can be positioned, monitored, and notified online in the form of goods to meet the e-commerce customers to check the real-time delivery information of the purchased goods at any time. Although this real-time information service has now been used in the shopping platform, it also can only provide node information and cannot see the transportation and distribution of the whole process and specific location, and the information will have different degrees of delay, so you can provide more accurate logistics information to improve customer satisfaction through big data, cloud computing, and other new generation of information technology.

4.3. Artificial Intelligence to Improve the Efficiency of Delivery. Lately, the improvement of computerized reasoning in China has become quicker and quicker, which has provoked the utilization of man-made consciousness calculations to turn out to be increasingly broad [26]. The broadness first calculation and A calculation are the principal ways of computing the ideal way for the conveyance of things and picking the course with a more limited way, which can make its conveyance benefits more self-evident (Figure 5). In the genuine conveyance period, on the off chance that the staff exclusively transport everything, the work proficiency will be poor and the expense of the work will be high. The utilization of astute robots and robots to convey merchandise will really work on the effectiveness and nature of the work and, furthermore, extraordinarily decrease the expense of the task. In the automated conveyance mode, on the off chance that it applies design acknowledgment and different advances, it can sweep and peruse a wide range of information data on the merchandise very rapidly, which will carry extraordinary comfort to the coordinated operations conveyance work. As the clients of online business operations are broadly appropriated and scattered, various kinds of clients should be coordinated with various sorts of conveyance modes. Under the savvy strategies, the conveyance mode can be additionally streamlined to give more decisions of conveyance modes, like home conveyance + conveyance time, self-administration pickup + self-decision pickup organization, pickup bureau + self-decision compartment, and so on, by clients as indicated by their own necessities, which are more adaptable conveyance activities to meet the separated requirements of web-based business clients. Along these lines, it can incredibly decrease the likelihood of dispatch conveyance disappointment, lessen the quantity of lost and harmed pieces, and work on the effectiveness of terminal conveyance.

4.4. Artificial Intelligence Changes the Function of Logistics Workers. Artificial intelligence has undoubtedly triggered the phenomenon of "machine for human" in the context of the new era of Industry 4.0 [27]. From a theoretical perspective, this is the substitution effect of artificial intelligence on workers' jobs. According to the PwC forecast (see Figure 6), as of 2030, for example, the United Kingdom may have 30% of jobs in automated production, lower than the United States and Germany, but higher than Japan, where the automation rate of warehousing, transportation, manufacturing, and wholesale and retail industries will obviously be higher. It is thus foreseeable that highly dangerous and repetitive manual labor and data collection jobs, including warehouse management delivery and distribution, will have the strongest replaceability.

With the transition phase of industrial transformation and upgrading, many transitional and prominent contradictions in job convergence cannot be ignored, such as structural unemployment caused by the mismatch between people and jobs. The extensive investment in "unmanned" equipment by logistics companies does not mean that human positions are not needed at all; on the contrary, many jobs still require more relevant human cooperation for better operation. In this context, society is bound to increase the number of positions adapted to the composite talent, that is,

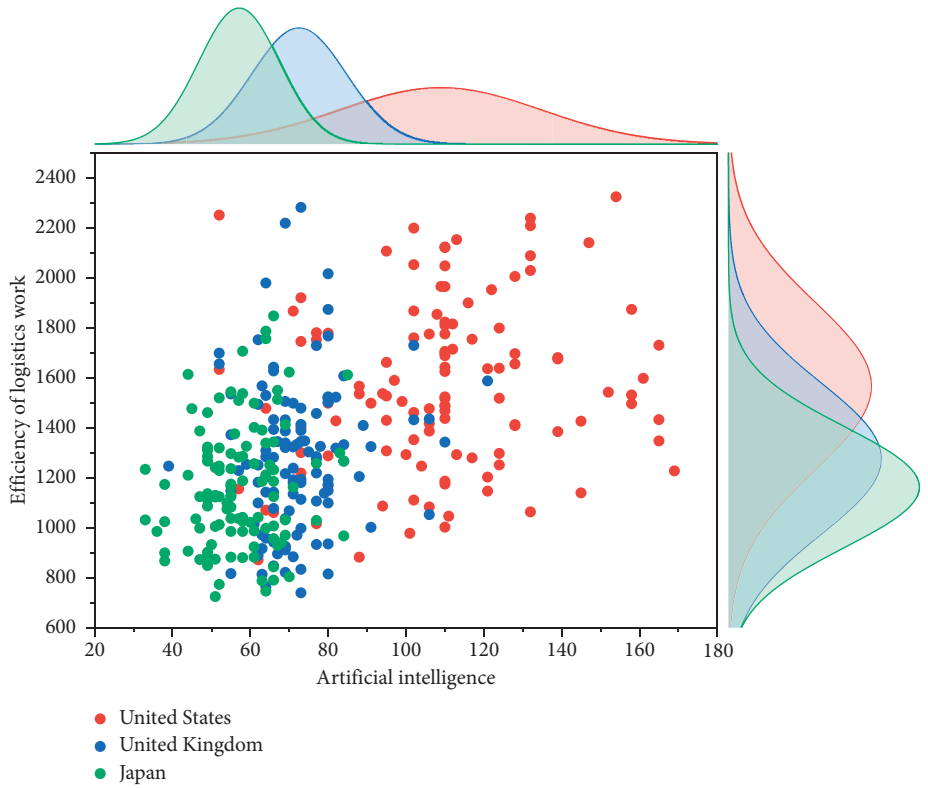


Figure 6: Matrix nephogram of the relationship between artificial intelligence with logistics work.

not only knowing how to operate the robot and familiar with the industrial process of high-quality personnel positions, which also means that those who cannot meet the future job skills demand of many front-line e-commerce warehouse employees will face the threat of unemployment.

In addition, this gives rise to the fact that development must face the "old" problem—the lack of professional composite high-quality talent. The rapid development of artificial intelligence technology is the result of the joint efforts of many data analysis experts, AI/machine learning senior engineers, data labeling professionals, AI hardware experts, and other professional talents, these mid- to high-end professional and technical personnel are definitely an important recruitment target for enterprises. Therefore, enterprises have stepped up the introduction of these market scarce resources to seize the opportunity in the industry competition. How to make the "old" problem in the "new" position can be solved makes artificial intelligence in the development of logistics enterprises face one of the key issues.

The advantage of artificial intelligence machines is that they can operate around the clock. At night, artificial intelligence machines are still running, making full use of the time that is difficult for natural people to use at night, and improving the efficiency of logistics [28]. But there are still many other nonstandard operations that cannot currently be transformed with AI. The most effective model may still be the human monitoring AI machines to collaborate to complete the logistics distribution so that AI machines become human eyes, legs, and hands to complete some fixed simple intelligent behavior, such as taking the planned distribution route, automatic identification of obstacles, and avoidance on the route. And people have to do something that cannot be predefined behavior, such as distribution of new routes, new user needs, and other tasks.

There is also the logistics of large items, the volume is small, and if the difficulty is too high with unmanned distribution, then only one person can complete it. The difficulty here includes the difficulty of realizing the loading tools for large items, and even if it is realized, the cost is still high due to the small amount. A feasible method is to use professional distribution units to implement large pieces of distribution, and then it is necessary to use the large pieces of distribution platform to realize the matching of demanders and providers so that it is both economic and fast; then, artificial intelligence can provide the accurate and timely help to do on-demand and on-time delivery to avoid waste. For example, the demander only needs to provide the photo of the bulky item to be delivered, the shipping location, and the receiving location. Based on visual recognition, artificial intelligence automatically matches the delivery tools and routes, as well as the related staffing. Fast speed and accurate matching are fully achievable. For the role played by people in logistics, under the artificial intelligence ecology, it should be monitoring and filling in the gaps. Do the delivery service that the machine cannot achieve and at the same time complete the machine's failure maintenance work. Do what the machine cannot do by itself.

Given the logistics of large items, the volume of logistics pieces is defined as a set $\chi = \{\mathbb{Z}_j\}_{j=1,2,\ldots,n}$, where J is

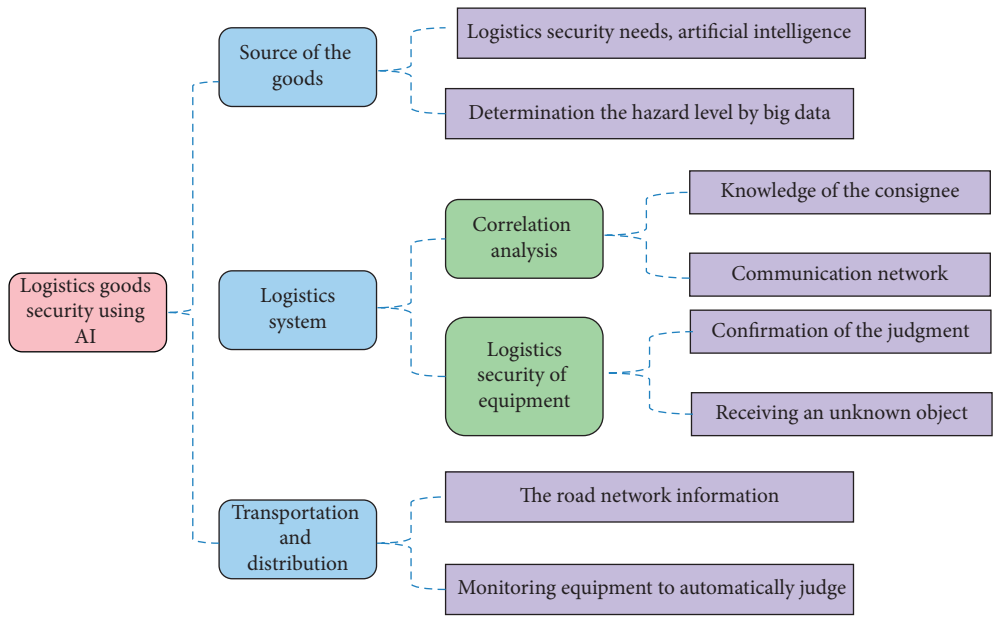


Figure 7: Frame diagram of artificial intelligence for logistics goods security prediction.

professional distribution units, and artificial intelligence system (u, v) of the jth logistics pieces in the demanders and providers is denoted by the vector $\mathbb{Z}_j \in x$. The loading tools for large items consist of on-demand and on-time delivery $\alpha_t(\mathbb{Z})$ at each shipping location providing confidence $\mathbb{R}_{jt} \in Rw \times h$ for each course j, where w and h are the delivery tools and routes, respectively, and t denotes the tth location. The first location of the corresponding photo of the bulky item uses visual recognition and artificial intelligence to provide confidence scores:

$$\begin{aligned} \alpha_t &= (\mathbb{Z}|\mathbb{N}), \\ \alpha_t &\subset \left\{s_1^j\left(\mathbb{Z}_j = \mathbb{Z}\right)\right\}, \\ \alpha_t &= (\mathbb{Z}|\mathbb{N}) \subset \left\{s_1^j\left(\mathbb{Z}_j = \mathbb{Z}\right)\right\}. \end{aligned} \tag{3}$$

All subsequent locations generate new confidence scores using the contextual information from the previous location:

$$\begin{aligned} &\alpha_t > [\mathbb{Z}|\mathbb{N}, \psi(\mathbb{Z}, \mathbb{R}_{t-1})], \\ &[\mathbb{Z}|\mathbb{N}, \psi(\mathbb{Z}, \mathbb{R}_{t-1})] \subset \left\{s_t^j\left(\mathbb{Z}_j = \mathbb{Z}\right)\right\}, \\ &\alpha_t > [\mathbb{Z}|\mathbb{N}, \psi(\mathbb{Z}, \mathbb{R}_{t-1})] \subset \left\{s_t^j\left(\mathbb{Z}_j = \mathbb{Z}\right)\right\}, \end{aligned} \tag{4}$$

where $\mathbb{R}_t \in Rw \times h \times (J+1)$ corresponds to the confidence score map of the related staffing of location t; ψ $(\mathbb{Z}, \mathbb{R}_{t\text{-}1})$ denotes the machine's failure maintenance work from the confidence map $\mathbb{R}_{t\text{-}1}$ to a wide range of pieces of distribution x.

4.5. Artificial Intelligence for Logistics Goods Security Prediction. Logistics goods security refers to, first, the security of the source of the goods themselves, such as using the logistics system to send dangerous goods to the victims; second, the security of the goods in the process of transportation and distribution, such as important goods, valuable goods, and flammable, explosive, and other dangerous goods [29, 30]. For the above two kinds of logistics security needs, artificial intelligence can help. The hazards at the source can be alerted by big data to determine the hazard level. Based on the correlation analysis between the destination of logistics goods and the place of issuance, a hazard alert is derived so that potential hazards can be avoided.

Generally speaking, logistics is always demand-pulled, and active delivery without demand is required to be reminded or ensured by relevant procedures. Some logistics may have active delivery; i.e., the consignee at the destination is shipped without the knowledge of the consignee. This can take advantage of the now developed communication network to first remind the consignee to confirm the judgment that the goods received are safe, which is always much safer than passively receiving an unknown object. Of course, some public sector services for the public and the establishment of the receiving department should have a special unpacking machine and testing equipment, through the safety test before the person confirms receipt. The second security, in fact, only needs to add video surveillance in each link of logistics and artificial intelligence monitoring equipment to automatically judge and interact with the big data of the road network information, to achieve early warning, such as abnormal road conditions ahead, congestion, or bad weather that damage roads. In short, we need to make full use of the potential of artificial intelligence to provide credible and feasible guarantees for logistics security. Figure 7 illustrates the frame diagram of artificial intelligence for logistics goods security prediction.

4.6. AI's Help for Logistics in Extraordinary Times (e.g., Under Public Epidemics). In extraordinary times, such as during plague outbreaks and earthquake disasters, logistics systems modified by artificial intelligence are good choices because of the restricted movement of people. Due to the plague caused

by a large number of people who cannot contact each other, then unmanned cars, unmanned logistics vehicles, and drones will bring convenience to the distribution of logistics so as to adapt to the logistics distribution in the extraordinary period. In addition, the waste caused by incorrect information about material distribution in the extraordinary period will be solved by the operation of artificial intelligence. The demand information will be summarized by big data, and then it will not be a problem to deliver the suitable materials to the suitable users.

5. Conclusion

With the development of information technology, the transformation to digitalization has become an important means to enhance the level of specialization in many industries. Cross-border e-commerce should cater to the development trend of "Internet+" and build a cross-border e-commerce logistics development mode based on "Internet+" by effectively integrating logistics links with the help of perception and identification and visualization digital technologies such as big data, Internet of Things, artificial intelligence, and 5G. It creates a comprehensive service platform with automatic perception and identification and visualization operation capability, promotes seamless docking of all links, realizes effective allocation of resources, and makes the operation cost of cross-border e-commerce logistics greatly reduced, and the operation level and service quality of cross-border e-commerce logistics are greatly improved. As an emerging technology, "artificial intelligence + logistics" has become the mainstream of the development of the logistics industry, improving all aspects of logistics. The utility of artificial intelligence in logistics is great, and it is vital for the development of the logistics industry. The rational use of artificial intelligence technology can realize the development of logistics industry changes, reducing the operating costs of logistics units at the same time, and can also better improve the efficiency of the work carried out. In the context of the rapid development of artificial intelligence technology, artificial intelligence in logistics operations has become the inevitable development of the times, through the use of the technology to better enhance the level of intelligence in logistics, the wisdom of logistics as the leading development of the logistics industry, abandoning the traditional solidified logistics model, to better respond to the development of the times, highlighting the advantages of artificial intelligence.

Data Availability

The datasets used during the current study can be obtained from the author upon reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] Z. Zhu, Y. Bai, W. Dai, D. Liu, and Y. Hu, "Quality of e-commerce agricultural products and the safety of the ecological environment of the origin based on 5G Internet of Things technology," *Environmental Technology & Innovation*, vol. 22, Article ID 101462, 2021.

[2] Z. Wang and H. Zhu, "Optimization of e-commerce logistics of marine economy by fuzzy algorithms," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 3813–3821, 2020.

[3] Y. Ding, M. Jin, S. Li, and D. Feng, "Smart logistics based on the internet of things technology: an overview," *International Journal of Logistics Research and Applications*, vol. 24, no. 4, pp. 323–345, 2021.

[4] S. Modgil, R. K. Singh, and C. Hannibal, "Artificial intelligence for supply chain resilience: learning from Covid-19," *International Journal of Logistics Management*, 2021.

[5] G. D. Konstantakopoulos, S. P. Gayialis, and E. P. Kechagias, "Vehicle routing problem and related algorithms for logistics distribution: a literature review and classification," *Operational research*, vol. 23, pp. 1–30, 2020.

[6] M. Nicola, Z. Alsafi, C. Sohrabi et al., "The socio-economic implications of the coronavirus pandemic (COVID-19): a review," *International Journal of Surgery*, vol. 78, pp. 185–193, 2020.

[7] W. M. Morrison, "China's economic rise: history, trends, challenges, and implications for the United States," *Current Politics and Economics of Northern and Western Asia*, vol. 28, no. 2/3, pp. 189–242, 2019.

[8] N. R. Mosteanu, A. Faccia, A. Ansari, M. D. Shamout, and F. Capitanio, "Sustainability integration in supply chain management through systematic literature review," *Calitatea*, vol. 21, no. 176, pp. 117–123, 2020.

[9] J. Huang, W. Yang, and Y. Tu, "Financing mode decision in a supply chain with financial constraint," *International Journal of Production Economics*, vol. 220, Article ID 107441, 2020.

[10] Z. Shen, "Research on the application of computer internet of things in intelligent logistics management," *Journal of Physics: Conference Series*, vol. 1648, no. 4, Article ID 042018, 2020.

[11] J. Wen, L. He, and F. Zhu, "Swarm robotics control and communications: imminent challenges for next generation smart logistics," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 102–107, 2018.

[12] G. Wu, "Research on the development path of logistics management innovation in e-commerce environment," *IOP Conference Series: earth and Environmental Science*, vol. 714, no. 4, Article ID 042022, 2021.

[13] S. S. Ali and R. Kaur, "Effectiveness of corporate social responsibility (CSR) in implementation of social sustainability in warehousing of developing countries: a hybrid approach," *Journal of Cleaner Production*, vol. 324, Article ID 129154, 2021.

[14] P. Dutta, S. Talaulikar, V. Xavier, and S. Kapoor, "Fostering reverse logistics in India by prominent barrier identification and strategy implementation to promote circular economy," *Journal of Cleaner Production*, vol. 294, Article ID 126241, 2021.

[15] S. Russell, "Human compatible: artificial intelligence and the problem of control," Penguin, Westminster, London, UK, 2019.

[16] S. Guo, X. Zhang, Y. Zheng, and Y. Du, "An autonomous path planning model for unmanned ships based on deep reinforcement learning," *Sensors*, vol. 20, no. 2, p. 426, 2020.

[17] X. Zhang, M. Zhou, H. Liu, and A. Hussain, "A cognitively inspired system Architecture for the mengshi cognitive vehicle," *Cognitive Computation*, vol. 12, no. 1, pp. 140–149, 2020.

[18] M. T. Gibbs, "Technology requirements, and social impacts of technology for at-scale coral reef restoration," *Technology in Society*, vol. 66, Article ID 101622, 2021.
[19] A. Lorenc and T. Lerher, "PickupSimulo–prototype of intelligent software to support warehouse managers decisions for product allocation problem," *Applied Sciences*, vol. 10, no. 23, p. 8683, 2020.
[20] D. Zhang, L. G. Pee, and L. Cui, "Artificial intelligence in E-commerce fulfillment: a case study of resource orchestration at Alibaba's Smart Warehouse," *International Journal of Information Management*, vol. 57, Article ID 102304, 2021.
[21] F. Wang, E. Lü, Y. Wang, G. Qiu, and H. Lu, "Efficient stereo visual simultaneous localization and mapping for an autonomous unmanned forklift in an unstructured warehouse," *Applied Sciences*, vol. 10, no. 2, p. 698, 2020.
[22] A. A. Ceder, "Urban mobility and public transport: future perspectives and review," *International Journal on the Unity of the Sciences*, vol. 25, no. 4, pp. 455–479, 2021.
[23] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, "Applications of artificial intelligence in transport: an overview," *Sustainability*, vol. 11, no. 1, p. 189, 2019.
[24] H. Fatemidokht, M. K. Rafsanjani, B. B. Gupta, and C. H. Hsu, "Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular ad hoc networks in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4757–4769, 2021.
[25] K. Zhang, S. Leng, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, "Artificial intelligence inspired transmission scheduling in cognitive vehicular communications and networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1987–1997, 2019.
[26] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, no. 1, pp. 195–199, 2019.
[27] M. Klumpp and H. Zijm, "Logistics innovation and social sustainability: how to prevent an artificial divide in human–computer interaction," *Journal of Business Logistics*, vol. 40, no. 3, pp. 265–278, 2019.
[28] M. Kohl, S. Knauer, and J. Fottner, "Industry 4.0 in logistics and associated employee competencies–a technology providers' perspective," *International Conference on Human Interaction and Emerging Technologies*, pp. 377–383, Springer, Cham, 2020.
[29] M. Bistron and Z. Piotrowski, "Artificial intelligence applications in military systems and their influence on sense of security of citizens," *Electronics*, vol. 10, no. 7, p. 871, 2021.
[30] T. S. Ramadoss, H. Alam, and R. Seeram, "Artificial intelligence and Internet of Things enabled circular economy," *International Journal of Engineering Science*, vol. 7, no. 9, pp. 55–63, 2018.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 4743216, 8 pages
https://doi.org/10.1155/2022/4743216

Research Article

Industry 4.0 Oriented Distributed Infographic Design

Lei He

Hefei Normal University, Hefei 230601, China

Correspondence should be addressed to Lei He; he1688@hfnu.edu.cn

Received 16 May 2022; Revised 16 June 2022; Accepted 21 June 2022; Published 25 August 2022

Academic Editor: Xingsi Xue

Since industry 4.0 was put forward in 2013, industrial process around the world has been moving rapidly towards the age of intelligent manufacturing. Industry 4.0 is known as the fourth industrial revolution dominated by intelligent manufacturing, which has changed the production mode of global manufacturing and triggered far-reaching industrial changes. However, when intelligent machines communicate with each other under industrial 4.0, a large amount of data adopting distributed control will be generated. The infographic in the data is mainly a visual design of industry 4.0 data. Therefore, this paper mainly studies the distributed data optimization processing for industry 4.0. Considering that data leakage is one of the biggest challenges faced by the data storage systems, this paper proposes a data storage method that considers the efficiency and security of data access. The concept of security distance not only guarantees data security but also takes into account the emphasis of different user groups on data security. To minimize data access time, this paper proposes a data access node selection algorithm to minimize data access time while ensuring data security. The simulation proves that compared with baselines, the data access time of the proposed algorithm in random topology and Internet2 topology is less than that of the current data storage algorithm while ensuring data security. The experimental results are simulated on Internet2 topology and random topology with Matlab and Omnet + + simulation platform, showing that the proposed algorithm can select the optimal data storage node under the condition of satisfying the security distance constraint, thus reducing the data access time.

1. Introduction

Industry 4.0 is proposed and applied at Hannover Messe in 2013, which is mainly aimed at the future manufacturing industry [1]. After the three industrial revolutions, it integrates network technology and digital technology to represent the fourth industrial revolution, which makes industry 4.0 attract high attention in the global industrial field [2, 3]. At present, industry 4.0 not only takes intelligent development as the primary target but also extensively applies advanced measures such as information technology, information interaction, and process reengineering. Based on meeting the personalized and differentiated needs of different consumers, flexible production is performed to achieve maximum decision optimization [4–6].

Today is the era of big data. In the era of Industry 4.0, the manufacturing industry will be built on an interactive platform based on the Internet and information technology. Industrial big data will become the core driving force of intelligent manufacturing [7, 8]. The main thinking direction of Industry 4.0 is to predict demand and production through data analysis and then use data to integrate the industry chain and value chain, so as to create greater value [9]. The production-related data are called the master data of the enterprise, which includes a series of product-related data such as design, process, modeling, test, maintenance, product structure, component configuration, and change records. These data are recorded, transmitted, and processed to enable the product to achieve life cycle management and further satisfy customers' personalized product needs [10, 11].

The huge amount of data in Industry 4.0 makes people suffer from information overload. In recent years, data analysis, data processing, and data presentation have become a research hotspot, among which infographics presented to users is a key link, which can enhance the readability and attractiveness of data information and increase the acceptance and dissemination [12]. Infographics is an excellent

way to present data and information concisely and clearly [13]. In the era of data and information explosion, higher requirements are put forward for the design of infographics, but how to present more rich data content from multiple perspectives more clearly and concisely has become a problem. The essence of infographic processing is a large amount of industrial data. The mass and diversity of industrial data make distributed systems become the best choice for data storage and management. Currently, distributed data storage systems are divided into peer-to-peer (P2P) storage technology and cloud storage system represented by cloud computing [14]. The advent of the big data era makes the research on the distributed storage system of great significance. For mass data storage, distributed data storage surpasses traditional centralized storage technology with its good scalability, robustness, and high efficiency.

Distributed data storage uses a large number of low-cost PC servers that are widely distributed in different geographical areas and connected to each other to store massive data [15]. This storage method can greatly save storage costs, but the availability of nodes is low. Meanwhile, the expansion of data storage greatly increases the probability of system failure. Based on cloud computing, cloud storage technology can combine different devices and different types of data to work together through application software, distributed file system, cluster technology, and network technology. However, storage nodes in different locations have different storage capabilities and link bandwidths, making it difficult to improve data access speed [16, 17]. In terms of data access time, graph partitioning is widely used at present [18]. This method has sufficient mathematical theory as support, but graph partitioning does not consider the node performance and link performance comprehensively, so it cannot solve the actual problem. How to reduce data access time while ensuring certain data security is the key point of distributed infographic design for industry 4.0.

To meet security requirements and support distributed infographic design of Industry 4.0, the concept of a K-distance topological subgraph is proposed in this paper; that is, in an undirected graph, if there is a subgraph whose distance between any two nodes is greater than K, then this subgraph is called the K-distance topological subgraph of the original graph. Based on the above definition, this paper uses K-distance topological subgraph in the original topology to place data so as to meet the security requirements. Moreover, to minimize data access time, this paper proposes a node selection algorithm based on a priority of nodes. The nodes are arranged in ascending order according to the access time of data, and then the data storage nodes are selected in turn under the constraints of security distance to form the optimal K-distance topological subgraph. Then, the data are placed on the K-distance topological subgraph.

Accordingly, the main contributions of this paper are summarized as follows:

(i) The concept of K-distance topological subgraph is proposed

(ii) A low complexity data placement algorithm is proposed

(iii) By comparing the effectiveness of the proposed algorithm on different network scales, the superiority of the proposed algorithm is proved

The rest of this paper is organized as follows. Section 2 reviews related work. In Section 3, we study the distributed data storage algorithms. The simulation results are presented in Section 4 and Section 5 concludes this paper.

2. Related Work

2.1. Data Analysis for Industry 4.0. Since industry 4.0 was put forward in 2013, the industrial process around the world has been moving rapidly towards the age of intelligent manufacturing. The development of data perception technology further helps to collect massive industrial data, and the innovation of industrial informatization is an opportunity. However, industrial data have the characteristics of large-scale, high-dimension, variable structure, and complex content, so it is a severe challenge to analyze industrial data. Diez et al. [19] conducted a comprehensive survey of the latest developments in data fusion and machine learning for industrial forecasting, focusing on identifying research trends, opportunities, and unexplored challenges. Peres et al. [20] proposed intelligent data analysis and real-time monitoring framework, which provided the basis for realizing scalable and flexible data analysis and real-time monitoring systems for the manufacturing environment. Raptis et al. [21] investigated the latest literature on the application of data management in a networked industrial environment and identified several open research challenges in the future. Costa et al. [22] aimed to find out the relationship or association between emerging technologies in industry 4.0 and applied data mining technology to a new bibliometric method to help identify association networks. Villalobos et al. [23] proposed a three-level hierarchical architecture for industrial 4.0 data storage in a cloud environment, which helped to manage and reduce the costs. Jiang et al. [24] proposed an analysis framework based on big data to analyze and extract the network behavior of cellular networks in industry 4.0 applications by using Hadoop and other technologies from the perspective of big data. Soltysik et al. [25] determined the trend and keywords for promoting the use of open data in industry 4.0. Li et al. [26] proposed a system framework based on the concept of industry 4.0, including the fault analysis and treatment process of machine center predictive maintenance.

2.2. Study for Distributed Data Storage. The large-scale use of the Internet has radically changed the data storage mode. With the increasing popularity of data sharing, local file systems cannot meet the needs of data sharing. More and more data are stored in distributed structures through the network. The distributed storage technology for file sharing emerges as the times require. Through the distributed data storage technology, people can easily and quickly exchange data and work together. Wu et al. [27] proposed a robust and auditable distributed data storage scheme to support safe and reliable edge storage in edge computing and ensure the

(i) Input: $G(V, E)$, and security distance K
(ii) Output: Nodes set K-dis-min-graph
(1) Select any node v
(2) Connect nodes with distance less than K
(3) $L1$: **for** $i = 1 \longrightarrow |adj(v)|$
(4) **for** $j = 1 \longrightarrow |adj(v_i)|$
(5) **if** $v_j \notin adj(v)$
(6) K-dis-min-graph $\leftarrow v_j$
(7) delete v
(8) $v = v_j$
(9) continue $L1$
(10) **end-if**
(11) **end-for**
(12) **end-for**
(13) return K-dis-min-graph

Algorithm 1: K-distance topology subgraph generation.

reliability and integrity of data in the distributed edge storage servers. Cangir et al. [28] preliminarily classified the blockchain-based distributed storage technology. Shi et al. [29] proposed a data placement algorithm based on fault-domain, which provided a new idea for the design of the distributed storage system. Yao et al. [30] introduced a remote image design of a dual node storage cluster, which could protect data in case of system failure. Liao et al. [31] considered a more practical data center network with fat-tree topology and used deep learning technology K-means to help store data blocks, so as to improve the read-write delay of data center networks. Jin et al. [32] introduced how to use distributed database HBase maintained by Apache to manage power data.

3. Distributed Data Storage Algorithm

3.1. K-Distance Topological Subgraph. Due to the limitation of security distance in data storage, it is necessary to find a list of storage node sets that meet the requirement of security distance before data chunks are placed [33]. To find such node sets, the concept of a K-distance topological subgraph is proposed in this paper.

Let $G(V, E)$ represent the network topology of a distributed storage system and be also an undirected connected simple graph, where V represents the set of storage nodes and E represents the link between the nodes. If there is a node set $V' \subseteq V$ and for $\forall v_1, v_2 \in V'$, $v_1 \neq v_2$, and we have dis_min $\geq K$, where *dis_min* represents the shortest hop number between two points, then V' is called the K-distance topological subgraph of graph $G(V, E)$.

Given the above, the K-distance topology subgraph V' of graph $G(V, E)$ is the set of nodes meeting the security distance limitation [34, 35]. Based on this, we propose a K-distance topology subgraph generation algorithm. The pseudo-code of Algorithm 1 is as follows.

According to Algorithm 1, given an undirected graph $G(V, E)$, select a node v arbitrarily at the beginning, then find the node v_j whose distance from this node is K, and then continue to find the point whose distance from v_j is K. Repeat this step until the graph G is traversed. The set K-dis-min-graph found is the topological subgraph of the K-distance.

According to the description of Algorithm 1, it is easy to get that the K-distance topology subgraph V' of graph $G(V, E)$ is not unique, as shown in Figure 1. Considering a 10-vertex topology graph $G(V, E)$, different initial nodes and intermediate nodes will be selected to obtain different K-distance topology subgraphs. Figure 1(b) is the schematic diagram of a 2-distance topology subgraph, and the node-set is {2, 4, 6, 9}. Figure 1(c) is also a 2-distance topological subgraph of graph $G(V, E)$ with a node-set of {1, 3, 5, 7, 8, 10}.

3.2. Storage Node Selection Algorithm. In this paper, the data placement problem satisfying certain security can be transformed into another problem; that is, given the security distance K, the problem of finding the K-distance topology subgraph satisfying the minimum data access time can be found in the network topology. As a result of the undirected graph, GK-distance topology subgraph is not unique, and this paper proposes an algorithm based on node priority, which arranges the nodes in order of unit data access speed. If the two nodes have the same access speed, they are arranged according to the node's self-protection capability (SPC). When selecting the storage node, the node with the highest priority should be selected as far as possible to ensure a high data access speed [36, 37]. Considering that the complexity of finding the K-distance topological subgraph is $O(n^2)$, a node selection algorithm is proposed in this paper, which minimizes the speed of data access and reduces the complexity of the algorithm based on satisfying the safe distance K.

SPC is the aggregate value of intrusion detection system capability value, anti-virus capability value, firewall capability value, and authentication mechanism capability value [38]. This paper assumes that all data center nodes have the above four security measures. Assuming that the data access point is node A in an undirected graph G, the unit data access speed from node v to data access point A is defined as for all nodes v in the graph. The pseudo-code of the data storage node selection algorithm (Algorithm 2) is as follows.

4. Simulation and Analysis

4.1. Simulation Environment. In this paper, Omnet++ [39] simulation platform and Matlab R2020a were used to verify the effectiveness of the storage node selection algorithm proposed in this paper. The network topology is divided into two types: random topology and Internet2 network connections [40], as shown in Figure 2. In Figure 2, the number on the line is the weight of the connection. In Figure 3, the larger the weight of the connection is, the thicker the connection line is. The former can measure the performance of data storage algorithms in various scenarios, and the latter can measure the performance of data storage algorithms in real scenarios. As shown in Table 1, we give the specific

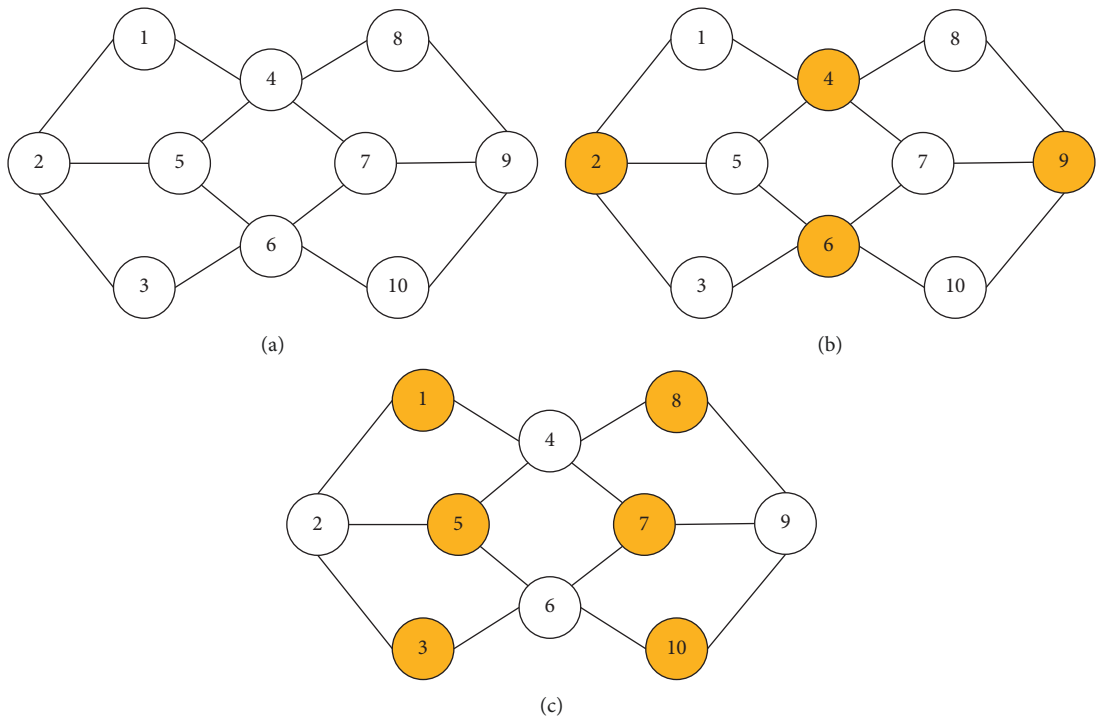


Figure 1: Schematic diagram of 2-distance topological subgraph. (a) Original network topology. (b) 2-distance topological subgraph 1. (c) 2-distance topological subgraph 2.

(i) Input: $G(V, E)$, K, Link bandwidth matrix, node A
(ii) Output: Optimal nodes set (Opt-nodes set)
(1) **for** $i = 1 \longrightarrow |V|$
(2) Unit data access speed = $\sum$ unit data/link bandwidth matrix
(3) **end-for**
(4) Rank the nodes according to step 2 from largest to smallest, and the ranked set is UDAS_D
(5) Opt_nodes set $\leftarrow$ UDAS_D_1
(6) delete UDAS_D_1 from UDAS_D
(7) **for** $i = 1 \longrightarrow$ |UDAS_D|
(8) dis = Dijkstra(A, UDAS_D_i)
(9) **if** dis $\geq K$
(10) Opt_nodes set $\leftarrow$ UDAS_D_i
(11) delete UDAS_D_i from UDAS_D
(12) **end-if**
(13) **end-for**
(14) return Opt_nodes set

Algorithm 2: Data storage node selection.

parameter and Table 2 shows settings of the simulation environment.

We compare the data access time of Algorithm 1 and Algorithm 2 proposed in this paper with that of CDPVDA [41], ACO-DPDGW [42], and UnifyDR [43].

(i) Cloud model-based Data Placement Algorithm with Virtual Data Agent (CDPVDA)

(ii) Ant colony optimization-based data placement of data-intensive geospatial workflow (ACO-DPDGW)

(iii) UnifyDR: A generic framework for unifying data and replica placement

4.2. Simulation results

4.2.1. Random topology. In this paper, we first compare the data access time results of various algorithms in random topological networks with different data volumes and network nodes, as shown in Figure 4. Figure 4 shows that with the increase of data volume, the data access time of the proposed algorithm is the smallest, which is about 50%

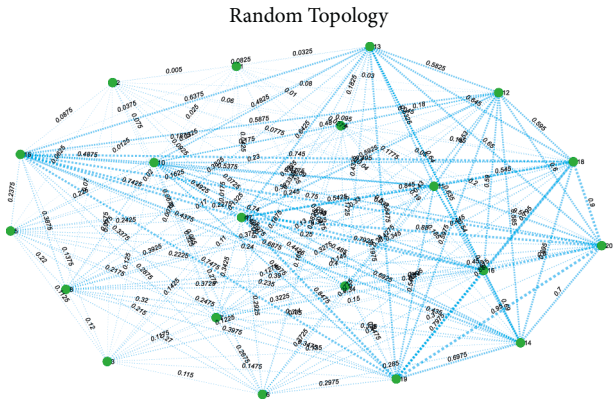


Figure 2: Topology of random network.

Figure 3: Topology of Internet2 network advanced layer 2 service.

Table 1: Simulation environment parameter settings of random network topology.

Parameter	Setting
Topological connection	Full connection
Number of nodes	20
Edge weight	Random
Edge style	Dotted line
Number of iterations	1000

Table 2: Simulation environment parameter settings of Internet2 network connections.

Parameter	Setting
Size of simulation area	4500 * 2700
Number of nodes	34
Capability of node storage (GB)	500–1000
Bandwidth (Mbps)	200–400
Data volume (GB)	2000–6000
Number of iterations	1000
Size of data chunks (MB)	100–500

shorter than that of baselines, and the data access time increases slowly. This is because the proposed algorithm adequately selects the nodes with good link condition to minimize the data access time. Figure 5 shows that with the increasing number of nodes, the data access time of the proposed algorithm is still the smallest compared with baselines, and the data access time is reduced by about 60%–

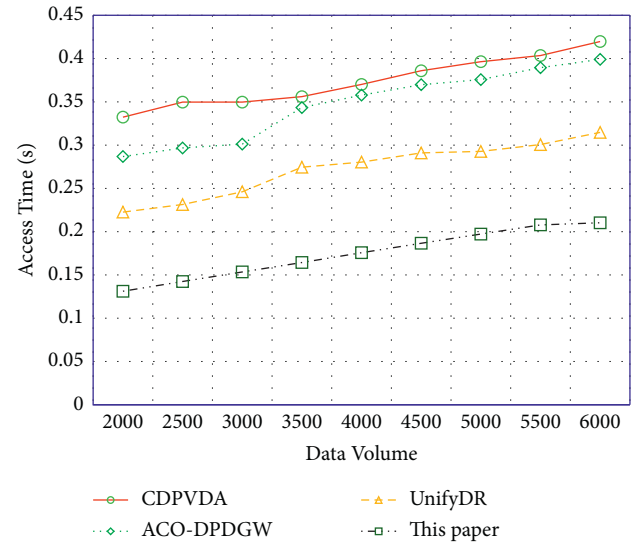


Figure 4: Data access time in random topology (different data volume).

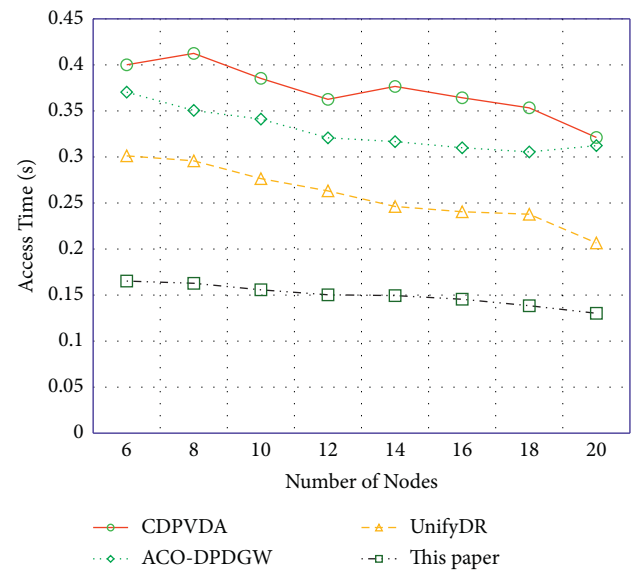


Figure 5: Data access time in random topology (different network size).

70% compared with baselines. This means that the proposed algorithm can select the best-performing nodes to store data under the condition of satisfying the security distance limit, thus minimizing the data access time.

4.2.2. Internet2 topology. As can be seen from Figure 6, as the volume of data in the Internet2 topology continues to increase, the data access time of all algorithms increases. As can be seen from Figure 6, data access speed on the Internet 2 topology is increasing with the increase of data volume, but the data access time of the algorithm proposed in this paper is still the smallest, and the data access time is reduced by about 50% compared with other baselines. Since the bandwidth in the Internet2 topology is 1 GBps, UnifyDR and

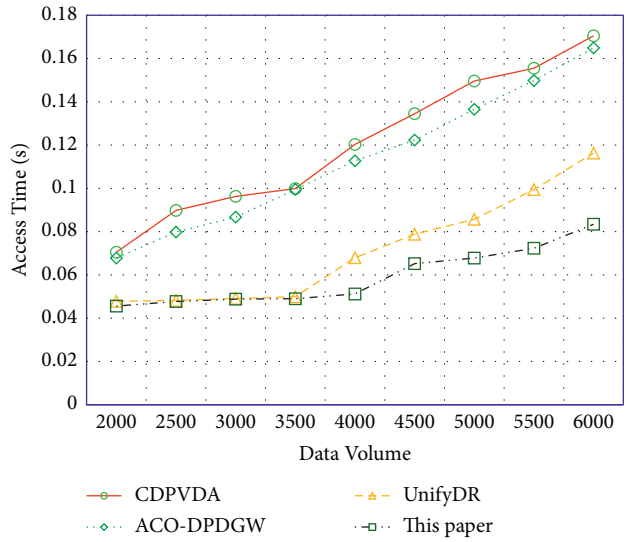


Figure 6: Data access time in Internet2 topology (different data volume).

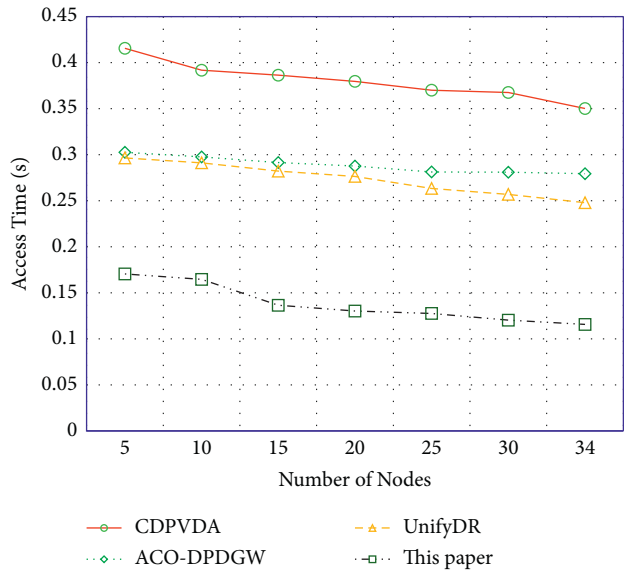


Figure 7: Data access time in Internet2 topology (different network size).

the algorithm proposed in this paper select the nearest nodes when the data volume is small, which makes that the data access time is the same. However, with the increasing of the data amount, when some nearest nodes are full of storage, the algorithm proposed in this paper is better than baselines in finding suboptimal nodes, so the performance of the algorithm proposed in this paper becomes better with the increase of data volume. As indicated in Figure 7, the data access time in the proposal is still reduced by about 50% compared with baselines with the increasing security distance.

5. Conclusions

As for the large amount of data generated by Industry 4.0, this paper proposes a data storage method considering the efficiency and security of data storage. Considering data security and user experience, to meet the needs of different user groups, the concept of security distance is proposed, which enables different users' requirements for data security to be used in the same data storage method. Considering the efficiency of data storage, a storage node selection algorithm is proposed to minimize data access time while ensuring certain data security, thus improving the user experience. Finally, simulation results show that compared with other existing data storage algorithms, the proposed algorithm can reduce data access time while ensuring certain data security.

In a distributed data storage system, a cloud storage system has a long distance between storage nodes and is generally distributed all over the world. However, a structured P2P network is highly volatile, which makes it difficult to ensure user experience in the networking strategy of the distributed data storage system. At present, the trend of distributed data storage system research is security, reliability, speed, and low energy consumption. Many existing works only optimize some of the above four conditions but do not achieve comprehensive optimization. Therefore, it is necessary to design distributed data storage methods and consistent maintenance policies that can meet the above requirements in the follow-up work to improve user experience.

Data Availability

All data used to support the findings of the study are included within this paper.

Conflicts of Interest

The author declares that there are no conflicts of interest in this paper.

Acknowledgments

This work was supported by Anhui Social Science Innovation and Development Research Project in 2021 (No. 2021cx136) and Anhui Quality Engineering Project in 2020 (No. 2020xfxm56).

References

[1] C. Bai, P. Dallasega, G. Orzes, and J. Sarkis, "Industry 4.0 technologies assessment: a sustainability perspective," *International Journal of Production Economics*, vol. 229, 2020.

[2] L. M. A. L. Dos Santos, M. B. da Costa, J. V. Kothe et al., "Industry 4.0 collaborative networks for industrial performance," *Journal of Manufacturing Technology Management*, vol. 32, no. 2, pp. 245–265, 2021.

[3] A. Schroeder, A. Z. Bigdeli, C. G. Zarcos, and T. Baines, "Capturing the benefits of industry 4.0: a business network

perspective," *Production Planning & Control*, vol. 30, no. 16, pp. 1305–1321, 2019.

[4] A. D. Neal, R. G. Sharpe, P. P. Conway, and A. A. West, "smaRTI-A cyber-physical intelligent container for industry 4.0 manufacturing," *Journal of Manufacturing Systems*, vol. 5, no. 2, pp. 63–75, 2019.

[5] R. I. Mogos, C. N. Bodea, M. I. Dascalu et al., "Technology enhanced learning for industry 4.0 engineering education," *Revue Roumaine des Sciences Techniques - Serie Électrotechnique et Énergétique*, vol. 63, no. 4, pp. 429–435, 2019.

[6] P. Ghadimi, C. Wang, M. K. Lim, and C. Heavey, "Intelligent sustainable supplier selection using multi-agent technology: theory and application for Industry 4.0 supply chains," *Computers & Industrial Engineering*, vol. 127, pp. 588–600, 2019.

[7] P. Basanta-Val, "An efficient industrial big-data engine," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1361–1369, 2018.

[8] D. Geng, C. Zhang, C. Xia, X. Xia, Q. Liu, and X. Fu, "Big data-based improved data acquisition and storage system for designing industrial data platform," *IEEE Access*, vol. 7, pp. 44574–44582, 2019.

[9] A. T. Keleko, B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne, *Artificial Intelligence and Real-Time Predictive Maintenance in Industry 4.0: A Bibliometric Analysis*, AI and Ethics, 2022.

[10] P. Radanliev, D. De Roure, K. Page et al., "Cyber risk at the edge: current and future trends on cyber risk analytics and artificial intelligence in the industrial internet of things and industry 4.0 supply chains," *Cybersecurity*, vol. 3, 2020.

[11] R. Kumar, S. P. Singh, and K. Lamba, "Sustainable robust layout using Big Data approach: a key towards industry 4.0," *Journal of Cleaner Production*, vol. 204, pp. 643–659, 2018.

[12] A. L. Harrison, N. F. Taylor, H. C. Frawley, and N. Shields, "A consumer co-created infographic improves short-term knowledge about physical activity and self-efficacy to exercise in women with gestational diabetes mellitus: a randomised trial," *Journal of Physiotherapy*, vol. 66, no. 4, pp. 243–248, 2020.

[13] W. Cui, J. Wang, H. Huang et al., "A mixed-initiative approach to reusing infographic charts," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 173–183, 2022.

[14] A. Tebbi, T. H. Chan, and C. W. Sung, "Multi-rack distributed data storage networks," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6072–6088, 2019.

[15] H. Wang, M. Xiao, C. Wu, and J. Zhang, "Distributed Classification for Imbalanced Big Data in Distributed Environments," *WIRELESS NETWORKS*, vol. 27, 2021.

[16] M. Al-Ruithe, E. Benkhelifa, Y. Jararweh, and C. Ghedira, "Addressing data governance in cloud storage: survey, techniques and trends," *Journal of Internet Technology*, vol. 19, no. 6, pp. 1763–1775, 2018.

[17] Y. Gupta, "Novel distributed load balancing algorithms in cloud storage," *Expert Systems with Applications*, vol. 86, p. 186, 2021.

[18] J. R. Nascimento, U. S. Souza, and J. L. Szwarcfiter, "Partitioning a graph into complementary subgraphs," *Graphs and Combinatorics*, vol. 37, no. 4, pp. 1311–1331, 2021.

[19] A. Diez-Olivan, J. Del Ser, D. Galar, and B. Sierra, "Data fusion and machine learning for industrial prognosis: trends and perspectives towards Industry 4.0," *Information Fusion*, vol. 50, pp. 92–111, 2019.

[20] R. S. Peres, A. D. Rocha, P. Leitao, and J. Barata, "Idarts - towards intelligent data analysis and real-time supervision for industry 4.0," *Computers in Industry*, vol. 101, pp. 138–146, 2018.

[21] T. P. Raptis, A. Passarella, and M. Conti, "Data management in industry 4.0: state of the art and open challenges," *IEEE Access*, vol. 7, pp. 97052–97093, 2019.

[22] M. B. Da Costa, L. M. A. L. Dos Santos, J. L. Schaefer, I. C. Baierle, and E. O. B. Nara, "Industry 4.0 technologies basic network identification," *Scientometrics*, vol. 121, no. 2, pp. 977–994, 2019.

[23] K. Villalobos, V. J. Ramirez-Duran, B. Diez, J. M. Blanco, A. Goni, and A. Illarramendi, "A three level hierarchical architecture for an efficient storage of industry 4.0 data," *Computers in Industry*, vol. 121, 2020.

[24] D. Jiang, Y. Wang, Z. Lv, S. Qi, and S. Singh, "Big data analysis based network behavior insight of cellular networks for industry 4.0 applications," *IEEE Transactions on Industrial Informatics*, vol. 162, no. 2, pp. 1310–1320, 2020.

[25] A. Soltysik-Piorunkiewicz and I. Zdonek, "How society 5.0 and industry 4.0 ideas shape the open data performance expectancy," *Sustainability*, vol. 13, no. 2, p. 917, 2021.

[26] Z. Li, Y. Wang, and K. Wang, "Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: industry 4.0 scenario," *Advances in Manufacturing*, vol. 5, pp. 377–387, 2017.

[27] J. Wu, Y. Li, F. Ren, and B. Yang, "Robust and auditable distributed data storage with scalability in edge computing," *Ad Hoc Networks*, vol. 117, 2021.

[28] O. F. Cangir, O. Cankur, and A. Ozsoy, "A taxonomy for Blockchain based distributed storage technologies," *Information Processing & Management*, vol. 58, no. 5, 2021.

[29] L. Shi, Z. Wang, and X. Li, "Novel data placement algorithm for distributed storage system based on fault-tolerant domain," *Journal of Shanghai Jiaotong University*, vol. 26, pp. 463–470, 2021.

[30] J. Yao, J. Shu, and W. Zheng, "Distributed storage cluster design for remote mirroring based on storage area network," *Journal of Computer Science and Technology*, vol. 22, pp. 521–526, 2007.

[31] Z. Liao, R. Zhang, S. He, D. Zeng, J. Wang, and H. Kim, "Deep learning-based data storage for low latency in data center networks," *IEEE Access*, vol. 7, pp. 26411–26417, 2019.

[32] J. Jin, A. Song, H. Gong et al., "Distributed storage system for electric power data based on Hbase," *Big Data Mining and Analytics*, vol. 1, no. 4, pp. 324–334, 2018.

[33] X. Xue and J. Chen, "Using compact evolutionary tabu search algorithm for matching sensor ontologies," *Swarm and Evolutionary Computation*, vol. 48, pp. 25–30, 2019.

[34] F. Zhu, J. Gao, J. Yang, and N. Ye, "Neighborhood linear discriminant analysis," *Pattern Recognition*, vol. 123, 2022.

[35] T. Wu, T. Wang, Y. Lee, W. Zheng, S. Kumari, and S. Kumar, "Improved Authenticated Key Agreement Scheme for Fog-Driven IoT Healthcare System," *SECURITY AND COMMUNICATION NETWORKS*, vol. 2021, 2021.

[36] W. Zhang, B. Zhang, W. Zhu et al., "Comprehensive Assessment of MODIS-Derived Near-Surface Air Temperature Using Wide Elevation-Spanned Measurements in China," *SCIENCE OF THE TOTAL ENVIRONMENT*, vol. 800, p. 800, 2020.

[37] X. Xue and C. Jiang, "Matching sensor ontologies with multi-context similarity measure and parallel compact differential evolution algorithm," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24570–24578, 2021.

[38] F. Zhu, Y. Ning, X. Chen, Y. Zhao, and Y. Gang, "On removing potential redundant constraints for SVOR learning," *Applied Soft Computing*, vol. 102, 2021.

[39] P. A. B. Bautista, L. F. Urquiza-Aguiar, L. L. Cardenas, and M. A. Igartua, "Large-scale simulations manager tool for OMNeT plus plus expediting simulations and post-processing analysis," *IEEE Access*, vol. 8, pp. 159291–159306, 2020.

[40] P. Kokoszka, H. Nguyen, H. Wang, and L. Yang, "Statistical and probabilistic analysis of interarrival and waiting times of Internet2 anomalies," *Statistical Methods and Applications*, vol. 29, no. 4, pp. 727–744, 2020.

[41] X. Zhang, Z. Hu, M. Zheng, J. Li, and L. Yang, "A novel cloud model based data placement strategy for data-intensive application in clouds," *Computers & Electrical Engineering*, vol. 77, pp. 445–456, 2019.

[42] X. Wu, Y. Liu, and C. Chen, "ACO-DPDGW: an ant colony optimization algorithm for data placement of data-intensive geospatial workflow," *EARTH SCIENCE INFORMATICS*, vol. 12, no. 4, pp. 641–658, 2019.

[43] A. Atrey, G. Van Seghbroeck, H. Mora, B. Volckaert, and F. De Turck, "UnifyDR: a generic framework for unifying data and replica placement," *IEEE Access*, vol. 8, pp. 216894–216910, 2020.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 6820073, 9 pages
https://doi.org/10.1155/2022/6820073

Research Article

A Reinforcement Learning-Based Basketball Player Activity Recognition Method Using Multisensors

Yang Bo

University of Shanghai for Science and Technology, Shanghai 200093, China

Correspondence should be addressed to Yang Bo; boyang021@usst.edu.cn

Received 16 May 2022; Revised 29 June 2022; Accepted 4 July 2022; Published 9 August 2022

Academic Editor: Xingsi Xue

It is an effective means to use a computer auxiliary system to assist athletes in training. In this paper, we design a technical activity recognition system for basketball players. The system uses the sensing module bound to the basketball player to collect the activity data and uses the proposed Multilayer Parallel Long Short Term Memory (MP-LSTM) algorithm to recognize the activity. Moreover, in order to extend the working time of the system and reduce the energy consumption of the sensing module, we also utilize the classical reinforcement learning algorithm DQN to adaptively control the sampling frequency of the sensing module for making a trade-off between recognition accuracy and energy consumption. Experiment results show that the recognition accuracy of the proposed MP-LSTM algorithm reaches 94%, while the recognition accuracy of the system remains at about 90% after applying the DQN algorithm, and the energy consumption is reduced by 76%.

1. Introduction

The basketball sport was invented in 1891 by an American physical education teacher named James Naismith. As a fun and easy event, this game is popular with the public, and it gradually spreads over the world [1]. Nowadays, basketball has become a world sport with detailed and rigorous rules, and it has higher requirements on the height, physical strength, and skill of the athletes. The National Basketball Association (NBA), which is the most professional league, has top-notch players in all aspects. The famous center, Shaquille O ′'Ne al, who is 214 cm tall, has a speed of 10.7 seconds for 100 meters and 47 seconds for 400 meters. Michael Jordan can reach 1.2 m in spot bounce as an outstanding basketball star. For the players, correct training methods are crucial to the improvement of performance on the court. The teaching ability and personal experience of the coach are very important to the athletes' training. However, with the continuous development of information technology, using computers and other intelligent devices to collect and analyze data can effectively assist athletes in the training [2].

The computer-aided training means are widely utilized in basketball sports. For example, Mieraisan [3] provides prototype implementations of computer vision algorithms in the sports industry, the main objective of his issue is to develop initial algorithms to solve play-field detection and player tracking in basketball game videos. Shah and Romijnders [4] use the deep learning algorithm to analyze the sports data of players, then predict whether a three-point shot is successful. Kizielewicz and Dobryakova [5] construct a kind of multicriteria decision-making method with an expert evaluation mechanism to analyze the players' data in order to accurately rank NBA players.

Inspired by related works, in this paper, we design a sports data acquisition and analysis system based on multisensors for basketball players. In the system, we use several homogeneous sensing modules to collect the technical motion data from basketball players and transmit the data to the server for analysis and processing. Meanwhile, considering that the battery power of the sensing module is limited and needs to be optimized for energy saving, we design an adaptive sampling frequency control algorithm based on reinforcement learning (RL). RL is one of the main methods in the field of machine learning and intelligent control in recent years. Its idea is to construct a strategy that enables the agent to choose behavior based on the environment in order to achieve the maximum cumulative reward. With the RL

means, the agent can choose the most appropriate behavior in different states. Therefore, RL is actually the learning of the mapping strategy, which is defined as "Policy," from environment state to agent behavior. The Deep Q-Network (DQN), a classic RL algorithm, involved in this paper was proposed by the Google DeepMind team in a paper published in Nature in 2015 [6]. The algorithm combines reinforcement learning and deep learning (DL), which makes it a major breakthrough in the field of artificial intelligence.

In this paper, we realize a system for data collection and analysis of basketball players based on multisensor architecture. In order to reduce the energy consumption of the sensing module, the system applies the DQN algorithm to adaptively control the sampling frequency of the nodes, so as to prolong the service cycle of the system. The rest of the paper is organized as follows: In Section 2, we introduce more related works on our issue, including information about the sensor-based recognition system and the DQN algorithm. Section 3 depicts the main structure of our system. Experiments are conducted in Section 4. Section 5 gives the conclusion of our work.

2. Related Works

2.1. Sensor-Based Activity Recognition for Basketball Players. With the maturity of computer technology, computer-based auxiliary training methods have become effective ways in professional sports training. The NBA is a highly professional basketball league, and some commercial organizations, e.g., Stats and Second Spectrum, record players' on-court performance with a multicamera system and provide professional data analysis services to the league teams. In order to personalize the training of players' technical movements and for the sake of flexibility, more studies choose to use an inertial sensing module to record and analyze activity data. Sangüesa et al. [7] establish a technical action data set using positioning sensors. The machine learning algorithm is used to recognize 5 classic basketball skill activities, including floppy offset, pick and roll, press break, post-up situation, and fast breaks, with an accuracy rate of 97.9%. Staunton et al. [8] use Magnetic, Angular Rate, and Gravity (MARG) sensors to measure the Counter-Movement Jump (CMJ) performance metric of elite basketball players. The experiment results show that there is a strong correlation between this index and players' competitive level. Mangiarotti et al. [9] design a wearable activity recognition system based on TinyDuino with an accelerometer and gyroscope, which is specially developed for coaches to track the activities of two or more players at the same time. Hasegawa et al. [10] study the wheelchair basketball game by adding inertia sensors to the athletes' wheelchairs to record the movement parameters so as to improve the athletes' skills in this game. Liu et al. [11] adopt a shapelet-based framework to recognize human activities and take daily life activities and basketball games to test the whole model. To summarize, given that sensing modules have a powerful function and wide popularity, it is an important means of using sensing devices to study basketball sports.

2.2. Energy-Saving Strategy. In many applications involving human activity recognition, it is usually a long-term monitoring job for the purpose of mining the subjects' activity patterns. However, the commonly used sensing modules are usually in small size and limited in power. Therefore, amount of studies try to minimize the energy consumption of the whole system and extend the work time of the system while keeping an acceptable recognition accuracy. There are lots of works on energy-saving strategies. For example, Phan [12] design a special algorithm for the use of GPS and turned on the GPS sensing module only when the geographical position of the user changed. The energy consumption of the entire activity recognition can be effectively reduced by utilizing the algorithm. Ling et al. [13] define the concept of "compression," the essence of which is to control the sampling frequency of the data acquisition module. Meanwhile, the paper also finds the optimal combination of sampling frequency and recognition accuracy through an exhaustive method. Wei et al. [14] reduce the sampling frequency of the sensing module to 2 Hz for energy saving. Meanwhile, the hybrid model structure of the decision tree + support vector machine is used to ensure the recognition accuracy. Gordon et al. [15] set the corresponding sensor configuration for different activities and predict the activity of the next window by referring to the history record. Due to the optimal sensor configuration, the recognition accuracy is improved and the power consumption is reduced. Morillo et al. [16] take advantage of different data sampling frequencies that vary from 32 Hz to 50 Hz to collect data and recognize daily activities (including walking, jumping, and cycling). According to the conclusion of these research studies, energy consumption is positively correlated with identification accuracy. However, all recognition systems are designed to achieve the highest accuracy with the least energy consumption, so making a balance between the two indicators is the core issue discussed in our work.

2.3. Reinforcement Learning. The concept of RL comes from the field of psychology, which is a branch of machine learning methods. RL algorithm can learn from interaction, which is similar to the evolution of the human learning process. The learning subject, named "Agent," learns knowledge according to the rewards or punishments obtained in the process for the purpose of adapting to the environment. The DQN is the combination of Q learning algorithm and deep learning. It takes advantage of the perception ability of deep learning to transform the state into a high-dimensional space and then utilizes the decision-making ability of Q learning method to map the high-dimensional state to a low-dimensional action space, thus solving the problem of dimension explosion. DQN is applied in areas such as game AI and resource optimization. Théate and Ernst [17] study trading algorithms in the stock market. They use DQN to determine the best trading timing and propose a more effective evaluation index for stock trading. The proposed approach can significantly improve both the safety and efficiency of online policy optimization based on

the simulation experiment results. Xu et al. [18] proposed a multiexit evacuation simulation based on Deep Reinforcement Learning (DRL) in the simulations on multiexit indoor, which is named as MultiExit-DRL. The proposed method presents great learning efficiency while reducing the total number of evacuation frames in all designed experiments. Sun et al. [19] utilize the DRL method to design a game system based on turn-based confrontation. In the model, they use a Q-learning algorithm to achieve intelligent decision-making. The experiments demonstrate the correctness of the proposed algorithm, and its performance surpasses the conventional DQN algorithm. Leng et al. [20] propose a Color-Histogram (CH) model, which combines the Markov decision process with a DQN algorithm, to solve the problem of color reordering in automotive spray painting workshops.

3. Methodology

3.1. Overall Structure. In this paper, we implement a basketball player skill activity recognition system based on a multisensor architecture. The overall structure is shown in Figure 1.

As shown in Figure 1, the main workflow of the recognition system is as follows:

(i) Data collection: the system collects motion data via sensing modules bound on the basketball player, including acceleration and angular velocity

(ii) Activity recognition: the preprocessed and segmented activity data are identified by the classifier

(iii) Recognition results: the recognition results are presented to the user and also passed to the sampling frequency controller as input parameters

(iv) Frequency control: the sampling frequency controller adaptively changes the frequency according to the recognition results to balance the power consumption and recognition accuracy

The main work of this paper lies in "Classifier" and "Sampling Frequency Controller," and details of both are given in the following parts.

3.2. Multilayer Parallel LSTM-Based Recognition Algorithm

3.2.1. The Structure of LSTM. The Long Short Term Memory (LSTM) is an upgrade structure of a classical deep learning model, the Recurrent Neural Network (RNN). RNN is a set of the model that can process time sequences and extract time-dependent features. In RNN, neurons in the same layer are connected with each other, and the latter neuron can use information from the former one. Therefore, in the process of network iteration, each neuron contains its own previous information, and its output is affected by the previous neurons, that is, the RNN network can remember and output time-dependent information of samples. However, due to gradient explosion or gradient disappearance occurs when processing long time series, in the traditional RNN network, the historical information that can be preserved by

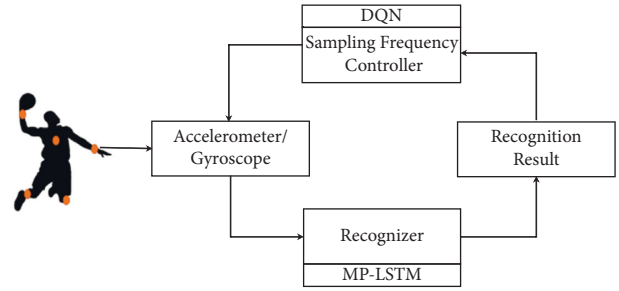


Figure 1: Overall structure of the system.

the hidden state is limited. The LSTM proposed by Hochreiter and Schmidhuber [21] for the first time in 1997 has become a classic solution to the two issues. The internal structure of the LSTM unit is shown in Figure 2.

LSTM is composed of 1 self-connected memory storage unit Ct and 3 gates that control memory information to be saved or forgotten. The Forget Gate determines what kind of information is allowed to pass through or kept. This structure ensures that errors will propagate in the network as a constant and prevents gradient explosion or gradient disappearance. It takes the current input X_t and the previous output state $h_t - 1$ of the hidden layer as input and calculates according to

$$f_t = \sigma\left(W_f X_t + U_f h_{t-1} + b_f\right), \tag{1}$$

where W_f is the weight matrix that maps the hidden layer input to the forget gate, U_f is the weight matrix that connects the output state of the previous moment to the forget gate, b_f is the bias vector, and σ is the activation function, which generally uses the sigmoid function.

The *Input Gate* determines whether the new information can be retained and updated to Ct for storage. The new information is controlled by X_t and h_{t-1}. A new state output s_t is obtained through the tanh function, then the input gate assigns a weight between 0 and 1 to each component of s_t to control how much new information is added to the network. The formulas are shown as follows:

$$\begin{aligned} i_t &= \sigma\left(W_i X_t + U_i h_{t-1} + b_i\right), \\ s_t &= \tanh\left(W_s X_t + U_s h_{t-1} + b_s\right). \end{aligned} \tag{2}$$

The output of the two gates is jointly calculated and updated to C_t after the information passes through the forget gate and the input gate:

$$C_t = f_t * C_{t-1} + i_t * s_t. \tag{3}$$

The *Output Gate* determines what information is output to the next unit. It uses X_t and $h_t - 1$ to calculate o_t, which is the state of the output information, and uses the tanh function to adjust the value of C_t to range [−1, 1].

$$\begin{aligned} o_t &= \sigma\left(W_o X_t + U_o h_{t-1} + b_o\right), \\ h_t &= o_t * \tanh\left(C_t\right). \end{aligned} \tag{4}$$

3.2.2. Multilayer Parallel LSTM. In this part, we introduce the multilayer parallel LSTM proposed in this paper, namely, MP-LSTM. Define an activity sequence $S = [S_1, S_2, \ldots, S_n]$, $S_i \in S$. The details in S_i are shown as follows:

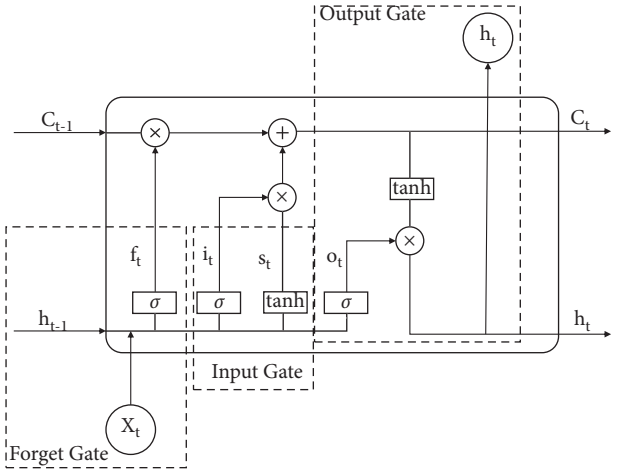


Figure 2: Structure of LSTM unit.

$$S_i = \{1 \le t \le m | [ACC\_X_t, ACC\_Y_t, ACC\_Z_t, GYO\_X_t, GYO\_Y_t, GYO\_Z_t]\}, \quad (5)$$

where ACC_X_t represents the X-axis acceleration data of the tth sampling point in S_i and GYO_X_t represents the angular velocity data. After the input sequence format is defined, the entire network structure is shown in Figure 3. Each input subsequence corresponds to an LSTM unit in the Parallel Layer, and each LSTM unit iterates the sample data in time order, so that the time-dependent information of each sample subfragment can be retained. Meanwhile, each LSTM unit has the same hyperparameters, such as a number of neurons and matrix shape, which ensures that each subfragment is processed in an equal manner.

In the Fuse Layer, there is only one LSTM unit that carries out the iterative calculation on the output matrix of the Parallel Layer and fuses the features of each sub-segment into a complete feature vector h. The elements in h are expanded and put into the Dense Layer for dimension reduction, so as to obtain the feature vectors of higher levels with lower dimensions of the sample. Finally, the softmax function is used to map the feature vector to the final probability results. The softmax function is as follows:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_c^C e^{z_c}}. \quad (6)$$

3.3. DQN-Based Adaptive Sampling Frequency Control Algorithm. When using wearable devices for activity recognition, changes of sampling rate affect the recognition accuracy and energy consumption, and the optimal sampling frequency is also different for activities. The algorithm proposed in this part carries out adaptive sampling frequency control on the sensing modules. It adjusts the sampling frequency according to the current activity, balances the recognition accuracy and energy consumption, and improves the overall performance of the system.

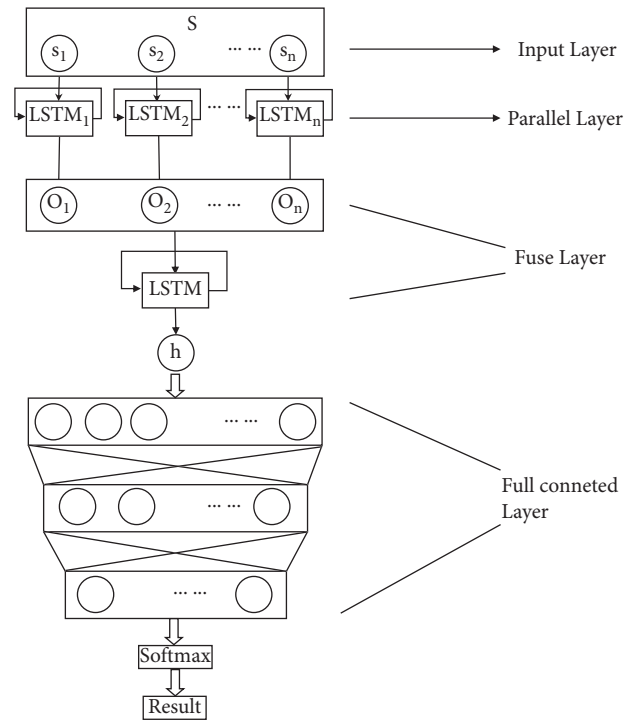


Figure 3: Structure of MP-LSTM.

3.3.1. Problem Mapping. Assume that the sensing module used can support k sampling rates, denoted as $F = (f_1, f_2, \ldots, f_k), f_1 < f_2 < \cdots < f_k$. The energy consumption of the module varies at different sampling rates. Let $P = (P_1, P_2, \ldots, P_k)$ denotes the sampling power set, the larger the sampling frequency f is, the larger the corresponding power consumption is, which means that $P_1 < P_2 < \cdots < P_k$. Suppose that there are m kinds of activities in the target activity set, denoted as $Y = (y_1, y_2, \ldots, y_m)$. Define a contiguous sample sequence with the label as Q, which is shown as follows:

$$Q = \{1 \le i \le n | (X_i, Y_i)\}, \quad (7)$$

where X_i is the activity data of the ith window in the sequence and Y_i is the corresponding activity label. We hope to design an activity recognition model to achieve high recognition accuracy, or a low recognition error rate, and meanwhile, the model has a sampling frequency selection strategy to minimize the overall energy consumption. The issue is depicted as follows:

$$\min_{\theta, f_1, \ldots, f_N} \sum_{t=1}^{N} l\{\hat{y}_t \ne y_t\} + \lambda \sum_{t=1}^{N} P_t, \quad (8)$$

where $\hat{y}_t = \underset{y}{\text{argmax}}\, p(y|x_t; \theta)$ is the recognition results, $l\{\hat{y}_t \ne y_t\}$represents the possibility that the output result is inconsistent with the actual label, and λ refers to the weight of energy consumption. For the convenience of calculation, $l\{\hat{y}_t \ne y_t\}$ can be replaced by the cross entropy of predicted probability distribution and behavior label, which is

$$l\{\widehat{y}_t \neq y_t\} = -\log\ pb(y_t|x_t;\theta), \tag{9}$$

where $pb(Y|X)$ is the probability that sample X belongs to category Y. Therefore, formula (8) can be changed as

$$\min_{\theta,f_1,\ldots,f_N} \sum_{t=1}^{N} -\log\ p(y_t|x_t;\theta) + \lambda \sum_{t=1}^{N} P_t. \tag{10}$$

3.3.2. Algorithm Description. In the traditional Q learning algorithm, the state space and the action space are usually discrete and finite, and the value function $Q(s,a)$ can be stored in tables. But in our work, the state space S is continuous and infinite, which means that the $Q(s,a)$ cannot be stored in tables. Therefore, DQN is used to train the model. The algorithm is shown in Figure 4.

The frequency controller adjusts the sampling rate according to the activity record. This process can be abstracted into a Markov decision process, so the sampling rate selector can be realized by RL. Here, we give the elements of the RL-based sampling frequency controller, including state space, action space, reward function, and action strategy:

(i) State space

The output activity probability distribution vector from the classifier is taken as the current state, which is

$$S = \{(pb_1, pb_2, \ldots, pb_i, \ldots, pb_m)|pb_i \in [0,1]\}, \tag{11}$$

where m is the number of activity types and pb_i is the probability that the sample label is i.

(ii) Action space

According to the previous assumption, different kinds of activities correspond to different sampling frequencies. The action space A is defined as follows:

$$A = (a_1, a_2, a_3, \ldots, a_i, \ldots, a_k), \tag{12}$$

where a_i means that the sampling frequency f_i is chose when collecting data with activity label i.

(iii) Reward function

Rewrite formula (10):

$$\begin{aligned}
&\min_{\theta,f_1,\ldots,f_N} \sum_{t=1}^{N} -\log\ p(y_t|x_t;\theta) + \lambda \sum_{t=1}^{N} P_t \\
&= \min_{\theta,f_1,\ldots,f_N} \sum_{t=1}^{N} -\log\ p(y_t|x_t;\theta) + \lambda P_t \\
&= -\max_{\theta,f_1,\ldots,f_N} \sum_{t=1}^{N} \log\ p(y_t|x_t;\theta) - \lambda P_t.
\end{aligned} \tag{13}$$

Then, the reward function is set as

$$R_t(s) = \log pb(y_t|x_t;\theta) - \lambda P_t. \tag{14}$$

Given that it is difficult to directly obtain the energy consumption value at different sampling rates in the experimental environment, we take a method of approximation instead. Assume that the energy consumption of a sensing device for single data sampling is fixed and denoted as P_0, then the power consumption of the sampling rate f_t is $P_t = f_t * P_0$. So the reward function is

$$R_t(s) = \log pb(y_t|x_t;\theta) - \lambda f_t P_0. \tag{15}$$

(iv) Action strategy

The ε-greedy is widely used in RL as an action strategy. It is an extension of the traditional greed mechanism. The agent chooses the action in the space A by

$$a = \underset{a\in A}{\operatorname{argmax}}\, Q(s,a). \tag{16}$$

The agent can select the actions randomly and freely under the ε-greedy mechanism. The specific strategies are as follows:

(1) Generating a random number in the [0, 1], i.e., num = Random(0, 1)

(2) If num $< \varepsilon$, then select the action according to the greedy mechanism, i.e., $a = \underset{a\in A}{\operatorname{argmax}}\, Q(s,a)$

(3) Else randomly select an action in A, i.e., a = Random(A)

ε is the greed degree and in the range of (0, 1). The agent may choose the random action easily if the ε is a small value, and the convergence speed gets slow correspondingly. Therefore, the choice of ε should be based on careful consideration.

We use pseudo code to demonstrate the DQN in Algorithm 1.

4. Experiment and Results

4.1. Dataset

4.1.1. Data Collection. We introduce the dataset used in the following experiments. We have 20 volunteers in the data collection, 10 of whom are basketball players and others are college students. Their physical information is given in Table 1. The whole activity dataset consists of 6 activities, including standing, standing dribble, penalty shot, jump shot, running, and running dribble. In order to unify these movements, we give a rigorous definition of these activities. For example, we define "standing dribble" as "Subjects should stand with their feet naturally apart, and try not to move, while keeping control of the basketball with one hand." These definitions can reduce the data diversity of the same activity.

Each subject is bound with 5 sensing modules, which are in the left thigh, right thigh, left leg, right leg, and torso in the data collecting process. The original sampling frequency is set to 50 Hz. According to experience from related work [22], we set the size of the sliding window as 3s, i.e., 150 data sampling points. Each subject performs each activity for 5 times, 3 min for each time.

4.1.2. Data Downsampling. In our work, we need to discuss the influence of different sampling frequencies on the recognition accuracy and energy consumption. The original

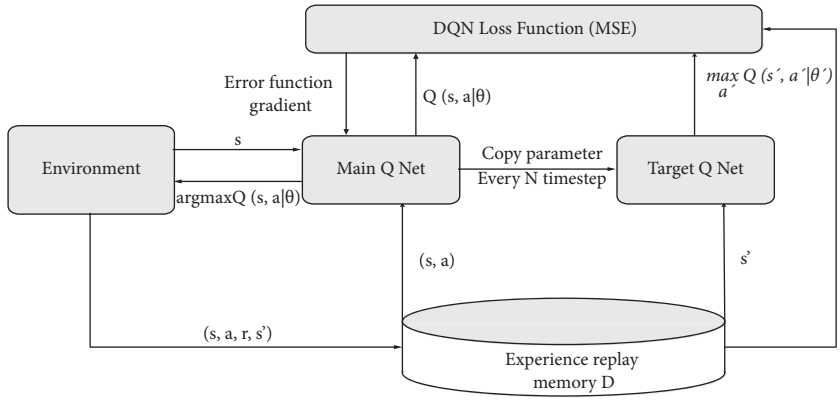


Figure 4: The DQN algorithm.

Init: D.size: $= N$, Main_Qnet.parameter: $= \theta$, Target_Qnet.parameter: $= \theta'$
For each episode:
  Init first sample $x_t = x_0$, Init first state $s_t = s_0$
  For each period:
    Choose the proper a_t according to the strategy ε, a_t: $= \underset{a}{\operatorname{argmax}} Q(s_t, a|\theta)$
    Execute a_t, update reward r_t, and next sample x_{t+1}
    Update state s_{t+1}
    Put experience sample (s_t, a_t, r_t, s_{t+1}) to D
    Random select experience sample (s_j, a_j, r_j, s_{j+1})
    If end_of_episode: $y_j = r_j$
    Else: $y_j = r_j + \gamma \max_{a'} Q(s_{j+1}, a'|\theta')$
    Update θ using loss function: $(y_j - Q(s_j, a_j|\theta))^2$
    Every C timestep: $\theta' = \theta$
  Until period end
End episode end

Algorithm 1: The DQN algorithm.

frequency is fixed (50 Hz), so it is necessary to conduct the downsampling to obtain data with different sampling frequencies. We downsample all the active data for 5 times to compose new datasets as given in Table 2.

4.2. Performance of MP-LSTM Recognition Algorithm. In this part, we test the recognition ability of the proposed MP-LSTM algorithm through experiments. We use the original data set described above for the recognition experiment with10 fold cross-validation. First, model parameters are set through experiments, including the number of LSTM units in the parallel layer, the number of hidden neurons in each LSTM unit, and the number of hidden neurons in the fuse layer. The experiment results are shown in Table 3.

The results above suggest that the increase in the number of LSTM units in the parallel layer improves the recognition accuracy. However, the accuracy rate is about 94.5%, which does not increase with more LSTM units. So we finally set this parameter as 6. The number of hidden neurons in each parallel LSTM affects the number of feature units sent to the next layer. Generally, a high feature dimension means powerful representation ability. However, according to the experiment, the best result is achieved when the number of hidden neurons is 24, and higher feature dimensions bring no significant promotion. Based on the same reason, we set the number of hidden neurons in the fusion layer as 64.

Table 1: Physical information of the volunteers.

Item	Value
Height	173.3 (±10.5)cm
Weight	68.2 (±9.7)kg
Age	26.5 (±3.5)
Sex	15 males : 5 females

We use the above parameters to build the MP-LSTM model and compare it with similar algorithms. Table 4 gives the comparison results.

According to the comparison results, the recognition accuracy of the algorithm proposed in this paper reaches 94.77%, which is better than other similar algorithms. Benefit from the parallel structure of the model, the recognition time is effectively reduced and the recognition efficiency is improved.

4.3. Performance of DQN Sampling Frequency Control Algorithm

4.3.1. Sampling Frequency vs Recognition Accuracy. In order to simplify the problem, we simplify the relation between frequency and energy consumption in the former parts. We

Table 2: Information of data downsampling.

Downsampling rate	New frequency (Hz)	Size of single sample (sample points)
2	25	75
4	12.5	38
5	10	30
8	6.3	19
10	5	15

Table 3: Relation between the MP-LSTM parameter and recognition accuracy.

No. of LSTM units in parallel layer	2	4	6	8	10
Recognition accuracy	93.88	94.19	94.57	94.00	92.67
No. of hidden neurons in parallel LSTMs	8	16	24	32	40
Recognition accuracy	89.14	93.08	94.33	93.91	93.85
No. of hidden neurons in fuse layer	16	32	48	64	80
Recognition accuracy	92.61	92.87	93.26	94.32	94.03

Table 4: Comparison with other algorithms.

Methods	Accuracy (%)
Dynamic time warping [23]	87.26
LSTM [24]	90.55
CNN [25]	90.78
PCA + SVM [26]	92.31
MP-LSTM	94.77

assume that the energy consumption is proportional to the sampling frequency. Therefore, the problem between energy consumption and recognition accuracy is transformed into making trade-off between sampling frequency and accuracy. We explore the influence of different sampling frequencies on the recognition accuracy in this part. Datasets of different sampling frequencies are recognized using the 10-fold cross-validation method, and the average recognition accuracy is given in Table 5. The $A1$–$A6$ in the table refers to the 6 activities, which are standing, standing dribble, penalty shot, jump shot, running, and running dribble.

It can be seen from Table 5 that the recognition accuracy of each activity decreases with the sampling frequency. However, the model has different performances on recognizing different activities. It achieves better results on standing, standing dribble, and free throw than on running and running dribble. It is supposed that the first 3 activities have smaller movements and concentrate on the upper limbs compare with the other 3 ones.

4.3.2. λ Determination. In this part, we realize the sampling frequency controller using the DQN algorithm, and the main parameter to be determined is λ. During training, 50 Hz is selected as the default frequency for the first sample of each sequence to ensure that enough features can be extracted. Meanwhile, we set the energy consumption of single sampling as $P_0 = 1$ to restrict the reward function in [0, 1].

After choosing training the sampling frequency controller, we evaluate the performance on the test set, focusing on the overall recognition accuracy and energy savings. Moreover, we define the Energy Saving Rate (ESR) to measure the energy saving effect. ESR refers to the percentage of energy saved by the currently selected sampling frequency compared to the original frequency. For example, the currently selected sampling frequency of 10 Hz provides an 80% ESR compared to the original sampling frequency of 50 Hz.

Figure 5 shows the relationship among recognition accuracy, ESR, and weight parameter λ. According to the figure, with the decrease of the weight parameters λ, the recognition accuracy continuously increases, while the ESR gradually reduces. This is because when λ is large, the model pays more attention to energy saving and tends to choose a lower sampling frequency, which also leads to a lower recognition accuracy. Correspondingly, when λ is small, the model pays more attention to the recognition accuracy, and the energy consumption is relatively high. In addition, when $\lambda > 0.5$, the accuracy rate increases obviously, while the energy saving rate decreases slightly, and vice versa. $\lambda = 0.5$ is an equilibrium state of the model, which makes the accuracy and ESR balanced. Therefore, λ is set to 0.5 in the subsequent experiments.

4.3.3. Overall Performance. Table 6 shows the recognition accuracy and ESR of each activity when $\lambda = 0.5$. According to the statistics, activities such as standing, standing dribble, and penalty shot have relatively high recognition accuracy compared with other 3 activities. Moreover, the ESRs of the first 3 activities overcome the values of A4–A6. We believe that the intensity of A1–A3 is lower and their data barely fluctuates, which means that data with a low sampling frequency is enough for recognizing these activities.

In addition, we compare the performance of the proposed model with other related works in Table 7. According to the results, the accuracy of the proposed model is lower than the methods in Reference [28], but with little difference. However, our model performs much better than the other 2 works in energy saving. To summarize, the proposed model can effectively reduce the energy consumption of the equipment, while ensuring a certain recognition accuracy.

Finally, in order to verify the performance of the DQN-based frequency control algorithm in the real scene, we

Table 5: Sampling frequency vs recognition accuracy.

Sampling frequency	Recognition accuracy (%)						
	*A*1	*A*2	*A*3	*A*4	*A*5	*A*6	Sum
50	98.92	97.80	95.75	91.41	93.30	89.37	94.43
25	98.51	97.12	95.44	91.01	92.91	88.44	93.91
12.5	98.10	95.31	94.26	89.90	90.08	86.27	92.32
10	96.25	94.87	91.78	88.52	89.29	83.37	90.68
6.3	90.71	87.52	87.22	86.08	86.27	80.56	86.39
5	88.27	83.15	85.74	84.70	83.75	79.46	84.18

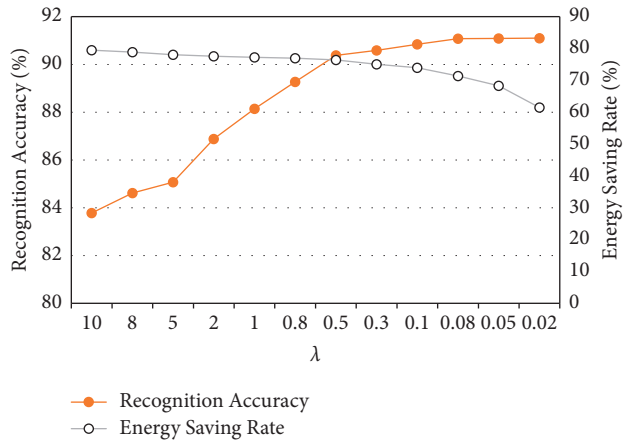


Figure 5: Influence on the recognition accuracy and ESR from different λ.

Table 6: RA vs ESR.

Item	*A*1 (%)	*A*2 (%)	*A*3 (%)	*A*4 (%)	*A*5 (%)	*A*6 (%)	Sum (%)
RA	96.25	95.84	93.21	89.46	85.69	81.83	90.38
ESR	79.20	80.71	79.94	77.52	72.10	68.81	76.38

RA: recognition accuracy; ESR: energy saving rate.

record the change of the remaining battery of sensoring module over time in 3 situations. Figure 6 gives the test results, where we have the following:

(i) Without DQN means to execute the recognition without DQN

(ii) DQN means to execute the recognition with DQN

(iii) NAN means no recognition

At the beginning of the experiment, the equipment battery is 100%. It takes 2.75 hours for the battery to decrease by 10% in "Without DQN," while the DQN algorithm slows down the speed of power decline by 39%. Compared with the "Without DQN" situation, the energy consumption of the sensing module is significantly reduced by 31%–39% in "DQN." At the end of these 2 situations, due to the fast speed of energy consumption in "Without DQN," the remaining power is 50% less than that in "DQN," indicating that DQN-based method is energy-efficient in the real scene.

Table 7: Compare with related works.

Related works	Recognition accuracy (%)	Energy saving rate (%)
[27]	89	50
[28]	92	28
Our work	90	76

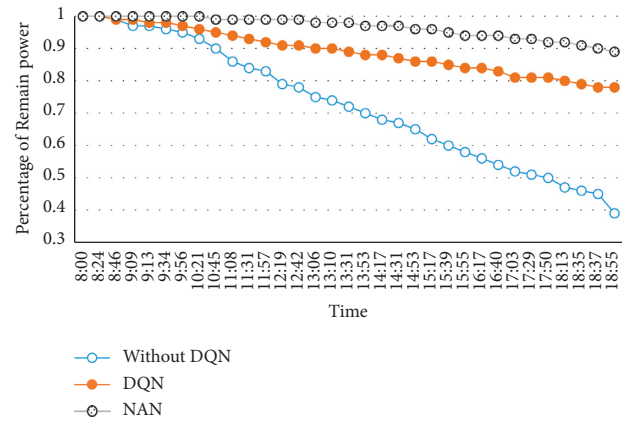


Figure 6: Test results under real scene.

5. Conclusions

In this paper, we implement an activity recognition system for basketball players using multisensor architecture. The MP-LSTM is utilized in the system for activity recognition, and the overall accuracy reaches 94.77%. Meanwhile, in order to prolong the working time of the sensing module, the DQN-based sampling frequency strategy is applied to adaptively control the sampling frequency of the module for energy saving. The experiment results show that the proposed method can reduce 76% of the energy consumption while maintaining the recognition accuracy at about 90%, which outperforms other related works.

The accuracy of activities such as jump shot, running, running dribble is relatively low. It is supposed that a great range of movements brings more noise to the activity data, which is a great challenge for the recognition model. In future work, we focus on this issue to improve the recognition ability of the model on activities with large movements.

Data Availability

All data used to support the findings of the study are included within this paper.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] J. Furey, "Basketball in America: a history," 2017, https://www.factmonster.com/sports/basketball/basketball-america-history.

[2] Y. J. Chang, "The use of computer-aided teaching software in the teaching of basketball tactics," *Applied Mechanics and Materials*, vol. 347-350, pp. 2777–2780, 2013.

[3] M. Mieraisan, *Basketball game analyzing based on computer vision*, Doctorial dissertation, Tampere University, Tampere, Finland, 2013.

[4] R. Shah and R. Romijnders, "Applying deep learning to basketball trajectories," *Computer Science: Neural and Evolutionary Computing*, vol. v1, 2016.

[5] B. Kizielewicz and L. Dobryakova, "MCDA based approach to sports players' evaluation under incomplete knowledge," *Procedia Computer Science*, vol. 176, pp. 3524–3535, 2020.

[6] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[7] A. A. Sangüesa, T. B. Moeslund, C. H. Bahnsen, and R. B. Iglesias, "Identifying basketball plays from sensor data; towards a low-cost automatic extraction of advanced statistics," in *Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 894–901, New Orleans, LA, USA, November 2017.

[8] C. A. Staunton, J. J. Stanger, W. Wundersitz, A. Gordon, E. Custovic, and I. Kingsley, "Criterion validity of a MARG sensor to assess countermovement jump performance in elite basketballers," *The Journal of Strength & Conditioning Research*, vol. 35, no. 3, pp. 797–803, 2021.

[9] M. Mangiarotti, F. Ferrise, S. Graziosi, F. Tamburrino, and M. Bordegoni, "A wearable device to detect in real-time bimanual gestures of basketball players during training sessions," *Journal of Computing and Information Science in Engineering*, vol. 19, no. 1, Article ID 11004, 2018.

[10] R. Hasegawa, A. Uchiyama, and T. Higashino, "Maneuver classification in wheelchair basketball using inertial sensors," in *Proceedings of the Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, pp. 1–6, IEEE, Kathmandu, Nepal, November 2019.

[11] L. Liu, Y. Peng, M. Liu, and Z. Huang, "Sensor-based human activity recognition system with a multilayered model using time series shapelets," *Knowledge-Based Systems*, vol. 90, pp. 138–152, 2015.

[12] T. Phan, "Intelligent energy-efficient triggering of geolocation fix acquisitions based on transitions between activity recognition states," in *Proceedings of the International Conference on Mobile Computing, Applications, and Services*, pp. 104–121, Springer International Publishing, Paris, France, November, 2013.

[13] X. Ling, M. Yu, and W. Kai, "Adaptive compressed classification for energy efficient activity recognition in wireless body sensor networks," in *Proceedings of the 2018 4th International Conference on Big Data Computing and Communications*, pp. 41–45, Chicago, IL, USA, August 2018.

[14] Z. Wei, Y. Yoshihara, D. Tang, N. Kubota, and N. Tay, "Energy-efficient activity recognition on smartphone," in *Proceedings of the Third International Conference on Computing Measurement Control & Sensor Network*, pp. 1–4, IEEE, Matsue, Japan, May 2016.

[15] D. Gordon, J. Czerny, T. Miyaki, and M. Beigl, "Energy-efficient activity recognition using prediction," in *Proceedings of the 2012 16th International Symposium on Wearable Computers*, pp. 29–36, Newcastle, UK, June 2012.

[16] L. Morillo, L. Gonzalez-Abril, J. Ramirez, and M. de la Concepcion, "Low energy physical activity recognition system on smartphones," *Sensors*, vol. 15, no. 3, pp. 5163–5196, 2015.

[17] T. Théate and D. Ernst, "An application of deep reinforcement learning to algorithmic trading," *Expert Systems with Applications*, vol. 173, no. 4, Article ID 114632, 2021.

[18] D. Xu, X. Huang, J. Mango, X. Li, and Z. Li, "Simulating multi-exit evacuation using deep reinforcement learning," *Machine Learning*, vol. 25, pp. 1542–1564, 2020.

[19] Y. Sun, B. Yuan, T. Zhang, B. Tang, W. Zheng, and X. Zhou, "Research and implementation of intelligent decision based on a priori knowledge and DQN algorithms in wargame environment," *Electronics*, vol. 9, no. 10, pp. 1–21, 2020.

[20] J. Leng, C. Jin, A. Vogl, and H. Liu, "Deep reinforcement learning for a color-batching resequencing problem," *Journal of Manufacturing Systems*, vol. 56, pp. 175–187, 2020.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] S. Bersch, D. Azzi, R. Khusainov, E. Achumba, and J. Ries, "Sensor data acquisition and processing parameters for human activity classification," *Sensors*, vol. 14, no. 3, pp. 4239–4270, 2014.

[23] S. Seto, W. Zhang, and Y. Zhou, "Multivariate time series classification using dynamic time warping template selection for human activity recognition," in *Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence*, pp. 1399–1406, Cape Town, South Africa, December 2015.

[24] Y. Chen, K. Zhong, Z. Ju, Q. Sun, and X. Zhao, "LSTM networks for mobile human activity recognition," in *Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications*, pp. 50–53, Bangkok, Thailand, January, 2016.

[25] C. A. Ronaoo and S. B. Cho, *Evaluation of Deep Convolutional Neural Network Architectures for Human Activity Recognition with Smartphone Sensors*, in *Proceedings of the Korea Information Science Society*, pp. 858–860, Montreal Canada, December 2015.

[26] C. A. Ronao and S. B. Cho, "Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models," *International Journal of Distributed Sensor Networks*, vol. 13, Article ID 155014771668368, 2017.

[27] Z. Yan, V. Subbaraju, C. Dipanjan, M. Archan, and A. Karl, "Energy-efficient continuous activity recognition on mobile phones: an activity-adaptive approach," in *Proceedings of the International Symposium on Wearable Computers*, pp. 17–24, IEEE Computer Society, Newcastle, UK, June 2012.

[28] V. Q. Viet, H. M. Thang, and D. Choi, "Adaptive energy-saving strategy for activity recognition on mobile phone," in *Proceedings of the IEEE International Symposium on Signal Processing & Information Technology*, pp. 95–100, IEEE, Ho Chi Minh City, Vietnam, December 2012.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 3179358, 9 pages
https://doi.org/10.1155/2022/3179358

Research Article

Multifactors Affecting Residential Well-Being in Urban Communities of Shenzhen Incorporating Intelligent Technologies

Xintong Wei,[1,2] **Guangtian Zou**,[1] **and Kin Wai Michael Siu**[2]

[1]*School of Architecture, Harbin Institute of Technology, Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin 150006, China*
[2]*School of Design and RISUD, The Hong Kong Polytechnic University, Hong Kong 999077, China*

Correspondence should be addressed to Guangtian Zou; zougt@hit.edu.cn

Received 7 June 2022; Accepted 2 July 2022; Published 30 July 2022

Academic Editor: Xingsi Xue

Rapid urban development is inseparable from technological advances, and the application of artificial intelligence in community life is becoming widespread, affecting residents' lifestyles and psychological well-being. This study investigated a variety of factors that affect the well-being of urban community residents. Environmental and emotional perceptions and overall well-being were assessed based on the responses of 179 respondents from six small communities in Shenzhen, China. Property management was strongly correlated with satisfaction with the physical environment and least correlated with neighborhood form. Pleasure, comfort, and the sense of belonging were correlated. In addition, interviews and questionnaires revealed a strong influence of AI facilities on people's well-being. Factor analysis revealed two-component matrices that explained more than 60% of the factors, which were described as "external" and "internal" factors. Finally, the study analyzes the relationship between intelligent devices and impact factors and their effects on residential well-being.

1. Introduction

1.1. Background. Since China's economic reform in 1978, the country has been developing at a rapid pace and various new technologies have been gradually applied in various industries. Urbanization has led to a significant increase in the number of people living in crowded spaces and illegal or informal settlements [1], creating a range of problems for people's lives and the environment [2]. At the same time, the application of artificial intelligence technologies in the building sector has contributed to the rapid development of smart buildings that optimize energy consumption and perform automatic adjustments while maximizing user comfort and satisfaction, and to some extent, the public health of residents [3]. Intelligent devices can greatly improve the efficiency of work through the effective integration of intelligent technology and the basic functions of the device, relying on the advantages of emerging technologies, and intelligent devices have been integrated into all aspects of people's lives. In urban communities, the use of intelligent access control systems, intelligent parking projects, intelligent homes, and intelligent alarm systems provides a diversity of intelligent services for people's lives.

There has been considerable evidence showing that people living in urban spaces in densely populated cities experience increased rates of stress and depression, and the living environment plays an important role in such situations. According to a report by WHO [4], an estimated 4.4% of the global population suffers from depressive disorder and 3.6% from anxiety disorder. In the context of high-density urban living, scholars have begun to study the importance of the environment in the improvement of urban living quality, psychological health, and well-being of urban residents [5, 6]. Well-being, which is related to the subjective feelings of individuals, is a complex concept in the field of psychology; SWB is a classical term that is widely researched. After SWB was first mentioned in the 1950s, its meaning was eventually described as a preponderance of positive over

negative affect [7]. From a subjective perspective, Andrews and Withey [8] found that most people consider SWB as an assessment of quality of life related to life satisfaction. With the development of research, Diener [9] proposed that SWB is a comprehensive judgement of individuals regarding their overall quality of life based on self-determined standards, referring to people's cognitive and affective evaluations of their own lives. He also stated that well-being is influenced by components of SWB, such as satisfaction with important domains, life satisfaction, low levels of negative affect, and high levels of positive affect [10]. This theory is also the theoretical support for the concept of "residential well-being" used in this study. However, most of the existing studies on well-being have been conducted from economic, cultural, social, and environmental aspects, and few scholars have considered the impact of intelligent devices on well-being. Based on these concepts and backgrounds, this study aims to explore the factors influencing residential well-being in urban communities in the context of the rapid development of smart technologies.

2. Potential Multifactors Affecting Residential Well-Being

To reach a deeper understanding of residential well-being, we reviewed the literature containing concepts relevant to our topic. As shown in Figure 1, mental or psychological well-being is influenced not only by individual characteristics or attributes but also by the socioeconomic circumstances in which people find themselves and the broader environment in which they live. Experts on emotions might argue that people's emotions, both positive and negative, evolved to help people assess their emotional state, and are therefore all equally desirable in appropriate circumstances; well-being experts assume that positive emotions are desirable and negative emotions are undesirable [9]. In this study, we consider residential well-being as a positive concept, as we focus on improving the urban living environment. The basic premise of this study is that residential well-being is a comprehensive concept, including the positive affect of residents and satisfaction with aspects of the environment.

Positive affect is defined as feelings and emotions that reflect a level of pleasurable engagement with the environment, such as happiness, joy, excitement, and contentment [11]. Emotions are a part of being human and are defined as psychological states brought about by subjective feelings [12]. In addition, there is some evidence that positive affect can facilitate behaviors reflecting a positive "approach" instead of social withdrawal [13]. From this perspective, experiences of positive affect could promote an individual's engagement with the built environment, which significantly impacts residential well-being. Additionally, some evidence suggests that people can improve their emotional well-being by cultivating experiences of positive emotions [14].

Given this background, in this study, we aim to assess the residential well-being in urban communities in Shenzhen, China. We used a questionnaire to evaluate residents' environmental and emotional perceptions, as well as overall

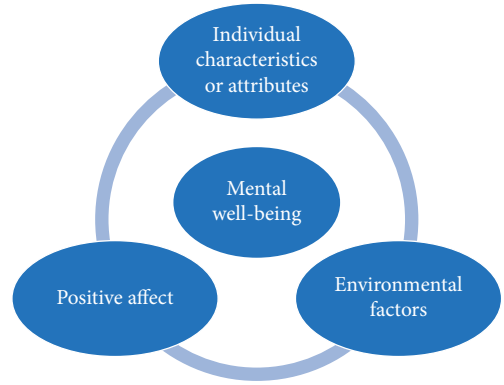


Figure 1: Multifactors of mental well-being.

residential well-being. By screening positive emotions related to the living environment, we finally summarized five emotions in the questionnaire, including the feelings of belonging, pleasure, security, convenience, and comfort. The five emotions reflect residents' positive affect, including engagement and positive emotions, assessed through self-report.

Regarding the environmental aspects, we collected residents' satisfaction with various elements in the living environment. According to reports from the WHO in 2018, neighborhood forms, housing quality, access to utilities and transport services, public green spaces, street safety, and social cohesion related to various aspects of the urban community may affect mental health to different degrees [15]. Additionally, many studies have suggested that some elements of the living environment can affect people's mental health or well-being, but these studies only focused on a specific aspect, which cannot cover most factors of the living environment. For example, neighborhood aesthetic quality and quantity of green spaces in living environments were demonstrated by a few studies to have positive associations with higher mental well-being [16, 17]; some other studies investigated the effects of changing the quality of housing on mental health and well-being in adults and the elderly individuals [18, 19]. Furthermore, Pollock et al. [20] proposed that the interaction between physical and mental health and changes in living environment, such as the distance from residence to public transportation or surrounding facilities, can be considered to affect mental health indirectly. Weinhardt et al. [21] explored that the experiences of public facility use are related to psychological well-being. According to the problems found in the field survey of the selected community and the face-to-face interviews with property management staff and residents, it can be known that property management has a great impact on residents' well-being, such as management of garbage classification and planning of parking spaces in the community. In this context, we summarize six elements, namely, "green space" and "neighborhood form" drawn from Bond and Gong [16, 17], "accessibility of transportation" drawn from Pollock [20], "public facilities in the community" drawn from Weinhardt [21], "property management" drawn from the

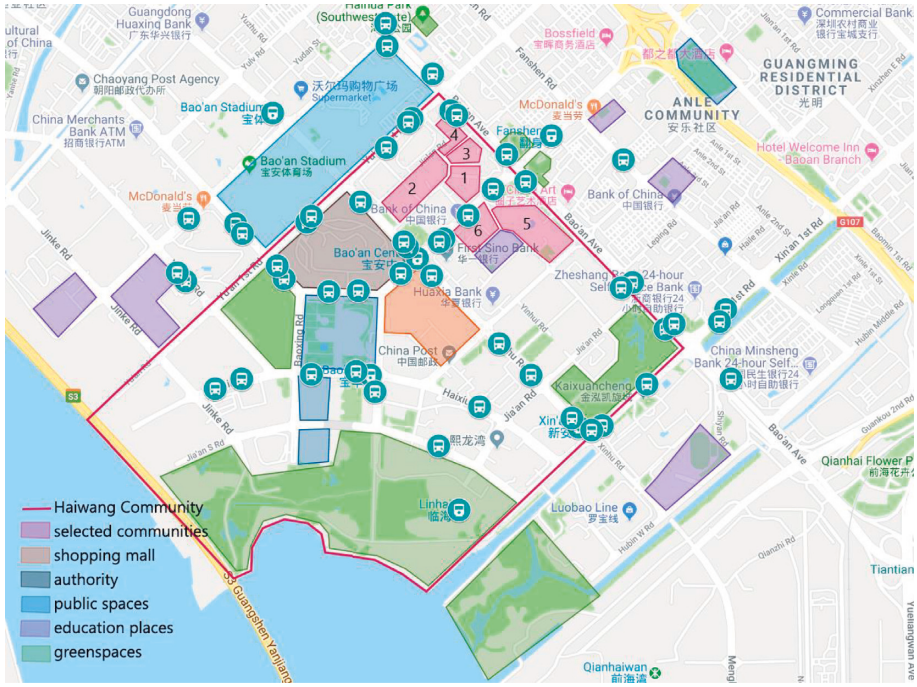


Figure 2: An overview of selected communities in Shenzhen.

findings of interviews, and "quality of housing" drawn from Rafaely and Tao [18, 19]. Instead of focusing on a single factor of the living environment, this study explored multiple environmental factors influencing residential well-being, especially emotional factors in relation to the psychological aspects. Therefore, this study proposes a hypothesis that six environmental factors and five affective factors all have an impact on residential well-being.

3. Methods

3.1. Selection of Study Location. To justify the above combination of factors, we selected several small communities in Shenzhen for our practical study. First, we chose Shenzhen (a densely populated city in China) as our target city because it has witnessed some of the most rapid and advanced infrastructure developments in recent years. In Shenzhen, AI technology is widely used in neighborhood construction, and there are many new small residential neighborhoods. By studying them, we hope to determine the human impact of the community environment under AI applications and explore ways to improve the process of continuous urbanization. Of the many small-scale residential communities in Shenzhen, Haiwang Community was an ideal candidate for our research. Haiwang Community (Figure 2) has a short history. It was built in 2004 and the construction of the community progressed very quickly. Since 2004, 11 high-rise commercial and residential communities have been built successively, most of which met our selection requirements, as detailed below.

First, the communities had a relatively organized property management system, which excluded some older small-scale communities. Second, the selection had a clear distinction between the internal and external areas of the community, to ensure that the study's subjects would understand the questions related to regional divisions in the questionnaire. Third, to avoid the influence of uncontrollable factors such as community culture and geographical location, we chose different small-scale communities in the same community to conduct the practical study. Fourth, we considered that the price of housing imposes certain restrictions on residents' personal economic conditions and some other related factors, such as social status and quality of life. Therefore, the price of housing within the selected residential areas was within a certain range (based on average house prices in Shenzhen) to ensure that the research results would be widely applicable.

Based on the above considerations, we chose six small-scale communities within Haiwang Community in the Baoan District of Shenzhen, which were coded as communities 1–6 (Figure 2) owing to ethical and privacy issues. Figure 2 shows an overview of the selected small-scale communities and their surroundings. There are 6 subway stations and more than 30 bus stops inside Haiwang Community and more than 10 other bus stops around the community. Additionally, there are many green spaces and public spaces inside and around the Haiwang Community, and a big shopping mall is located near the selected small-scale communities. Outside Haiwang Community, there are some sports clubs and Boan Stadium, which are only a 15-minute walk from the selected small-scale communities.

3.2. Questionnaire Design. As the purpose of this study was to enable residents to report what made them feel good about their living environment, instead of letting researchers define residential well-being for them, a questionnaire was considered the most effective and efficient way of capturing

individuals' experience of residential well-being [22]. In addition, although self-reporting of global SWB and life satisfaction might be influenced to some extent by transient factors, a considerable number of evidence shows that SWB is a stable state, and measures of it show considerable temporal reliability [23]. All questions about satisfaction levels and feelings about the living environment in the first section of the questionnaire were closed-ended questions because they are easier for respondents to answer and can reduce the number of irrelevant or confusing answers as well [24].

The questionnaire covered four aspects of self-assessment: basic information, overall residential well-being, satisfaction with environmental factors, and perceptions of different emotions. This study aimed to reveal the multifactors of residential well-being in urban communities. However, when people were asked about a particular emotion or environmental aspect, they would answer the first thing that came to mind and neglect, why they answered in that specific way and whether it was accurate. To avoid preconceived impressions, respondents were first asked about their overall residential well-being instead of their perception of a single factor. To investigate the relationship between residential well-being and environmental aspects, respondents were required to assess their level of satisfaction in relation to different environmental factors.

Considering the perception of emotions, when people were directly asked about a specific emotion, they might have had a different understanding of its meaning, which could have affected the validity of the research results to some extent. Therefore, to avoid misunderstandings, respondents were asked to evaluate the degree of their agreement with different descriptions, instead of rating the intensity of their emotions. At the end of the questionnaire, respondents were asked to rate their expected well-being considering the smart devices they were currently using and assuming that multiple smart devices would be introduced into their residential community in the future.

Respondents' answers were evaluated on a 5-point Likert-type scale. They were asked to rate the following aspects:

Overall residential well-being, from 1 (lowest) to 5 (highest).

Six items regarding environmental satisfaction (green space, neighborhood form, public transport, open space design, property management, and housing quality), from 1 (totally dissatisfied) to 5 (totally satisfied).

Five statements about emotions ("I feel that it is very convenient to travel and live here," "The layout and facilities here make me feel comfortable," "The management here makes me feel safe," and "Living here I feel a part of the community," and "The public environment and architectural appearance in the community make me feel happy"), from 1 (totally disagree) to 5 (totally agree).

3.3. Data Collection. The following approaches were applied in this study to collect data. The first approach was to set up an online questionnaire and send it to residents' online contact groups, which was convenient for the respondents. To ensure the authenticity of the collected data, we included a question in the online questionnaire for screening unqualified respondents who did not live in the selected communities. In addition, the online questionnaire had a submission limit so that each account could submit the questionnaire only once. The second approach was to randomly distribute questionnaires in-person, on weekday evenings in the public areas of the community, which was more efficient for the researchers. Moreover, the questionnaires were distributed randomly on different floors of each building to improve the validity of the study. In the adoption of these two approaches, we collected 201 questionnaires. Following this, we screened the questionnaires, excluding those that were not answered completely or had obvious problems. After the screening, there were 179 valid questionnaires left.

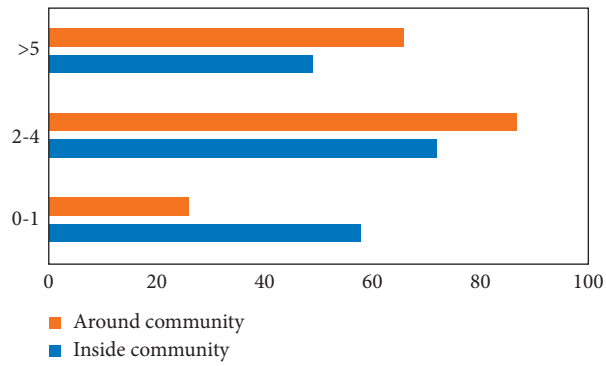


Figure 3: Frequency of use of exercise spaces inside and around the community.

Table 1: Residential well-being level of five communities.

	N	Mean	Minimum	Maximum
Community 1	43	3.40	1	5
Community 2	28	3.79	3	5
Community 3	32	3.56	2	5
Community 4	37	3.51	2	5
Community 6	30	4.00	1	5
Total	170	3.62	1	5

3.4. Data Analysis. SPSS 21.0 was used to establish a database containing all results of both online and on-site questionnaires, including multifactors [12]. The data were analyzed using the following methods: a statistical analysis to understand the basic data of respondents, a correlational analysis to calculate the relationships between different factors and residential well-being, and the factor analysis of emotional and environmental responses to extract factors that could summarize the results of the residential well-being questionnaire.

4. Results and Analysis

4.1. Statistical Analysis of Field Data. The gender balance of the respondents was 43% men and 57% women. The

Table 2: Correlations between environmental satisfaction and residential well-being.

		Green space satisfaction	Neighborhood form satisfaction	Public transportation satisfaction
Residential well-being	Correlation	0.582	0.332	0.377
	Significance (2-tailed)	0.000	0.000	0.000
		Open space design satisfaction	Property management satisfaction	Housing quality satisfaction
	Correlation	0.599	0.620	0.553
	Significance (2-tailed)	0.000	0.000	0.000

education level of the respondents was mainly junior college/under graduation, with 76.5% of respondents being undergraduates. Respondents were mainly between 15 and 59 years of age (66.5% were aged 35–59 years and 31.3% were aged 15–34 years). In terms of family structure, 90.5% of respondents lived with direct relatives, while 5.6% of respondents lived on their own. With respect to the distribution of respondents based on the duration of their stay in the community, the number of residents who lived there for more than 5 years was the highest, accounting for 63.7%, and the number of respondents who lived between 1–5 years accounted for 23.5%.

In addition, we interviewed the respondents about the frequency of their use of exercise spaces inside and around the community to understand more about their daily lives. As shown in Figure 3, 67.6% of respondents often exercised (2–4, 5, or more than 5 times weekly) inside the community public spaces, while 32.4% of respondents rarely exercised inside the community public spaces. Of the total respondents, 85.5% respondents exercised in areas around the community. Overall, the majority of residents living in Haiwang Community like to carry out activities in the living environment, and therefore, we consider that the results of these respondents can reflect the residents' perceptions of the living environment to some extent.

As listed in Table 1, we compared six small-scale communities in the Haiwang Community to ascertain residential well-being based on residents' self-reports. Since the data from community 5 were from less than 30 respondents (because of some uncontrollable reasons), we excluded the respective communities when considering the overall residential well-being in each community. As listed in Table 1, the overall residential well-being in Haiwang Community was 3.62. Although community 6 had the highest overall sense of well-being, the majority of residents living in Community 2 felt a higher sense of residential well-being because no one in Community 2 gave an answer of less than 3 points.

4.2. Correlations between Different Factors and Residential Well-Being. Correlation analysis was performed by combing the data of six small-scale communities, to find the associations between residential well-being and satisfaction with different environmental aspects. Table 2 lists the results of environmental satisfaction and residential well-being levels, with the community as the control variable. All six aspects were significantly positively correlated with residential well-being. While the satisfaction with property management seemed to have the most significant correlation with residential well-being, the satisfaction with green space and open space design also showed significant correlations with residential well-being in the result. Although the neighborhood form showed the weakest association with residential well-being, it still reached a correlation value of 0.332. Public transportation showed the second-lowest significance in the correlation results. It seemed that space design and housing quality had more positive correlations with residential well-being than did public transportation.

Table 3 lists the results of correlation analysis between five emotions and residential well-being based on respondents' self-assessment, with the community as the control variable. The results illustrated that five emotional factors had a significant positive correlation with residential well-being. Among them, the sense of pleasure was the most closely related to residential well-being (reaching 0.661). While the sense of convenience showed the lowest correlation in comparison, however, it was still significantly correlated with residential well-being. In addition, comfort and belonging played essential roles in the results, with correlations of 0.531 and 0.516, respectively. The results showed that positive psychological feelings such as pleasure and comfort showed more significant correlations with residential well-being than practical aspects such as convenience and security.

4.3. Relationship between Emotional Factors and Environmental Satisfaction. We also conducted a correlation analysis between emotional factors and environmental satisfaction to explore the relationship between different aspects of factors and check whether the factors could be interpreted using principal component analysis. As listed in Table 4, each of the six factors of environmental satisfaction was significantly correlated with each of the five emotional factors, respectively, which were suitable for principal component analysis. Green space had a significant correlation with respondents' pleasure level, which was also the highest correlation in the results. Additionally, the pleasure level was clearly correlated with open space design and property management. While public transportation significantly impacted the convenience level (correlation of 0.589), housing quality showed a high correlation with the comfort level (0.537). Although neighborhood form did not have the

Table 3: Correlations between emotional factors and residential well-being.

		Convenience level	Comfort level	Security level	Belonging level	Pleasure level
Residential well-being	Correlation	0.344	0.531	0.461	0.516	0.661
	Significance (2-tailed)	0.000	0.000	0.000	0.000	0.000

Table 4: Relationships between emotional factors and environmental satisfaction.

		Convenience level	Comfort level	Security level	Belonging level	Pleasure level
Green space satisfaction	Correlation	0.291	0.579	0.398	0.542	0.702
	Significance (2-tailed)	0.000	0.000	0.000	0.000	0.000
Neighborhood form satisfaction	Correlation	0.182	0.341	0.255	0.391	0.309
	Significance (2-tailed)	0.015	0.000	0.001	0.000	0.000
Public transportation satisfaction	Correlation	0.589	0.419	0.371	0.423	0.430
	Significance (2-tailed)	0.000	0.000	0.000	0.000	0.000
Open space design satisfaction	Correlation	0.249	0.595	0.388	0.547	0.692
	Significance (2-tailed)	0.001	0.000	0.000	0.000	0.000
Property management satisfaction	Correlation	0.283	0.578	0.386	0.487	0.693
	Significance (2-tailed)	0.000	0.000	0.000	0.000	0.000
Housing quality satisfaction	Correlation	0.431	0.537	0.437	0.517	0.607
	Significance (2-tailed)	0.000	0.000	0.000	0.000	0.000

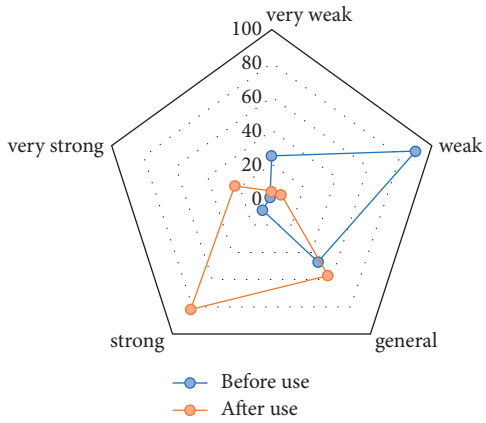


Figure 4: Comparison of the residential well-being before and after using intelligent devices.

strongest association with residential well-being, the results showed a strong correlation between neighborhood form and five emotion levels.

4.4. Comparison of the Residential Well-Being before and after Using Intelligent Devices. As shown in Figure 4, we compared the two data from the questionnaire on overall living satisfaction. The results show that when people consider the use of intelligent devices or think about the integration of intelligent devices into their lives in the future, there is an overall upward trend in their residential well-being. A significant increase in the frequency of strong and very strong residential well-being can be found with the use of smart devices. This reveals that the introduction of intelligent devices can compensate to a certain extent for the negative impact of other environmental factors on residents and indicates that residents' acceptance of intelligent devices is positive and eager.

Table 5: Rotated component matrixes.

	1	2
Green space satisfaction	0.815	0.184
Neighborhood satisfaction	0.395	0.260
Public transportation satisfaction	0.249	0.785
Open space design satisfaction	0.862	0.128
Management satisfaction	0.827	0.158
Housing quality satisfaction	0.644	0.415
Convenience level	0.098	0.882
Comfort level	0.747	0.292
Security level	0.471	0.509
Belonging level	0.647	0.464
Pleasure level	0.846	0.273

4.5. Factor Analysis of Residential Well-Being. Since there were significant correlations among all factors, we conducted factor analysis to explore the deep relationship and coefficient matrix of different factors. The result of Kaiser–Meyer–Olkin and Bartlett's test was 0.909, which meant the data could be analyzed through factor analysis. Based on the results of the principal component analysis, we selected the top two components to explain 63.56% of the 11 factors. Table 5 lists the results of the rotated component matrixes. Component 2 was strongly correlated with public transportation satisfaction, convenience level, and security level, which were always related to the surrounding conditions and geographical location of the urban communities. Thus, we named Component 2 an "external factor." The other eight factors had strong correlations with Component 1. We summarized Component 1 as "internal factor" for all related factors showing significant correlations with the internal conditions and psychological feelings.

5. Discussions

This study investigated the effects of environmental satisfaction and emotional factors on residential well-being in

urban communities in Shenzhen. A practical field study was carried out by collecting questionnaires and interviewing residents, where respondents evaluated their overall residential well-being, satisfaction with different environmental aspects, and levels of five emotional factors. The three main findings are discussed below.

First, the results showed that residential well-being was significantly associated with six environmental aspects and five emotional responses. Satisfaction with neighborhood and public transportation had the weakest correlation with residential well-being, as reflected by the results of the correlation analysis. This finding corresponded to the result that the feeling of convenience showed the lowest correlation with residential well-being. This suggests that residential well-being had more to do with the psychological situation than with the convenience of transportation. Additionally, the selected small-scale communities in this study were in the same community, so the transportation conditions and convenience levels of the living environment were similar for the respondents. Based on the interviews of some respondents, we were able to understand that when the residents were asked about their sense of residential well-being, they may think more about their feelings toward other factors and neglect the convenience of transportation to some extent. In terms of the neighborhood form, the reason for the low correlation could be the change in lifestyle in modern cities like Shenzhen. When people move to high-rise buildings, they progressively have less communication with their neighbors and might not even know the people who live next door. Interviews revealed that the application of intelligent facilities in neighborhoods allows people to live more efficiently and independently but at the same time reduces the opportunities for neighborhood communication. For example, many new neighborhoods in Shenzhen have applied face recognition technology to the systems of neighborhood gates and unit doors (Figure 5), which enhances the convenience of living in the neighborhood while greatly reducing the socialization and mutual assistance between neighbors. As a result, the alienation of the neighborhood was considered unimportant when respondents evaluated their residential well-being. In other words, the factors related to transportation and neighborhood form showed the least correlation with residential well-being in this study.

We also found that property management was the most strongly associated factor with residential well-being in terms of environmental satisfaction. This finding was not expected before analyzing the data. The reason might be the overall improvement of community construction. When the basic facilities of the living environment were satisfied, people began to pursue better levels of living conditions and factors such as property management and service reflected the quality of the community to some extent. According to the interviews, most respondents mentioned the problems of garbage disposal and sanitation, which were within the scope of property management. Another reason might be the importance placed on garbage sorting in Shenzhen during the time the field study was conducted, as the local authority encouraged recycling. During the time the study was conducted, considerable information about garbage sorting

Figure 5: Face recognition to enter the unit building (taken by the author).

policies was displayed publicly in the community. This context made people more aware of property management and regardless of whether this created a good or bad impression, it might have influenced the results of this study as it was easily recalled when respondents responded to the questionnaire. The above reasons are suggested to demonstrate why property management showed the highest correlation with residential well-being in the results. In addition, open space and green space played an important role in the environmental satisfaction results, which reflects residents' attention to the physical environment. In terms of emotional factors, the results showed that feelings of pleasure and comfort had the highest correlations with residential well-being, followed by feelings of belonging. According to these results, when talking about residential well-being, people considered and cared more about positive perceptions of the living environment such as pleasure, comfort, and belonging than the practical aspects such as whether the environment was safe and convenient. This finding was also consistent with the description of SWB, which is more connected to people's emotions and feelings. In addition, although this study found that the use of intelligent devices can enhance the residential well-being to a certain extent, what in-depth links exist between intelligent devices and the factors influencing residential well-being still need to be further explored.

5.1. Application of Artificial Intelligence in Enhancing Residential Well-Being. Based on the results of factor analysis, we can summarize two main components of the 11 influential factors: external factors and internal factors. External factors included levels of convenience, security, and satisfaction with public transport, which were mainly related to public transport and physical facilities around the community. On the other hand, internal factors included satisfaction with green space, open space, property management, neighborhood form, housing quality, and levels of comfort, belonging, and pleasure. These eight factors were primarily related to the physical environment and management inside the community, and therefore, we called these components internal factors. During the interview process, some residents had strong expectations for the intelligence of the neighborhood, and existing studies have shown the positive impact of an intelligent built environment on human well-being. Therefore, we explore the

Figure 6: Community management cloud platform (taken from Yunmou Community and Xingtianxia Applet).

Figure 7: Infrared alignment detectors (retrieved from http://www.youboy.com/s504637858.html).

possible effective applications of AI technology in enhancing residential well-being.

First, the application of the community management cloud platform (Figure 6) contains the integration of various intelligent technologies such as face recognition, password door opening, and smart door lock. This not only promotes communication efficiency between property and residents and facilitates timely access to community information, but the use of the smart platform also greatly reduces the burden of travel for residents by eliminating the need for traditional key and door cards. In addition, the platform's display of community activities may increase the likelihood of residents' social engagement, thus improving the sense of belonging in the community. In enhancing community security, the diverse applications of infrared sensors help improve the overall alarm system of a neighborhood, as shown in Figure 7. When it is applied to the monitoring system, the alarm signal is connected to the control center through the network, and when an intruder enters the community boundary, an alarm signal is sent. The application of infrared sensor technology helps reduce crime rates and enhance residents' sense of security while facilitating property and security personnel to manage neighborhood security. Infrared sensors can also be applied to electrical equipment and fire prevention, enhancing the quality of housing while safeguarding the lives and property of residents.

Although Shenzhen was selected as a representative sample as a typical city with the rapid urbanization of technology, our focus on Shenzhen only would lead to limitations in the study findings. Therefore, future research should use more diverse data and analysis methods to investigate in depth the mechanisms of built environment influence on residential well-being and what kind of impact artificial intelligence can have in promoting residential well-being, and further develop intelligent built environment factors related to residential well-being. By investigating which environmental factors can positively affect residents' emotions, community environment design strategies to improve well-being can be developed in the future through the adaptation of intelligent technologies to improve the residential well-being of urban communities.

6. Conclusions

This study explores the factors influencing residential well-being and the impact of smart devices in residential communities on well-being. Six environmental factors and five affective factors were found to be significantly associated with residential well-being. Factors related to transportation and neighborhood form have the least association with residential well-being, while property management is the factor most strongly associated with residential well-being. In terms of the emotional factors, pleasure and comfort were most associated with residential well-being. In addition, it was found that the application of intelligent devices can enhance residential well-being, but the specific influence mechanism still needs further study.

Data Availability

The data are available from the first author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Harbin Institute of Technology and The Hong Kong Polytechnic University. The research was funded by the National Natural Science Foundation of China (52078158). The authors would also like to acknowledge the support of Eric C. Yim Endowed Professorship in Inclusive Design (8.73.09.847K).

References

[1] World Health Organization, *Our Planet, Our Health: Report of the WHO Commission on Health and Environment*, WHO, Geneva, Switcherland, 1992.

[2] X. J. Yang, "China's rapid urbanization," *Science*, vol. 342, no. 6156, p. 310, 2013.

[3] K. Cui and C. Liu, "Artificial intelligence + elderly services: development model and realization path," *Chinese Journal of Gerontology*, vol. 42, no. 8, pp. 2037–2044, 2022.

[4] World Health Organization, "Depression and other common mental disorders," 2017, https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf.

[5] S.-C. C. Lung, C.-R. Lee, and S.-C. Hu, "Inequality of Asian-type neighborhood environmental quality in communities with different urbanization levels," *Environmental Science & Policy*, vol. 38, pp. 1–10, 2014.

[6] G. Sansom, J. Parras, A. Parras et al., "The impacts of exposure to environmental risk on physical and mental health in a small geographic community in Houston, TX," *Journal of Community Health*, vol. 42, no. 4, pp. 813–818, 2017.

[7] A. Cronin-de-Chavez, S. Islam, and R. R. C. McEachan, "Not a level playing field: a qualitative study exploring structural, community and individual determinants of greenspace use amongst low-income multi-ethnic families," *Health & Place*, vol. 56, pp. 118–126, 2019.

[8] N. M. Bradburn, *The Structure of Psychological Well-Being*, Aldine, Chicago, IL, USA, 1969.

[9] F. Andrews and S. Withey, *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*, Plenum Press, New York, NY, USA, 1976.

[10] E. Diener, "Subjective well-being," *Psychological Bulletin*, vol. 95, no. 3, pp. 542–575, 1984.

[11] L. A. Clark, D. Watson, and J. Leeka, "Diurnal variation in the positive affects," *Motivation and Emotion*, vol. 13, no. 3, pp. 205–234, 1989.

[12] C. Peterson, *A Primer in Positive Psychology*, Oxford University Press, Oxford, UK, 2006.

[13] J. T. Cacioppo, W. L. Gardner, and G. G. Berntson, "The affect system has parallel and integrative processing components: form follows function," *Journal of Personality and Social Psychology*, vol. 76, no. 5, pp. 839–855, 1999.

[14] B. L. Fredrickson, "Cultivating positive emotions to optimize health and well-being," *Prevention & Treatment*, vol. 3, no. 1, 2000.

[15] World Health Organization, "WHO housing and health guidelines," 2018, https://apps.who.int/iris/bitstream/handle/10665/276001/9789241550376-eng.pdf?ua=1.

[16] L. Bond, M. Egan, A. Kearns, and C. Tannahill, "GoWell: the challenges of evaluating regeneration as a population health intervention," *Preventive Medicine*, vol. 57, no. 6, pp. 941–947, 2013.

[17] Y. Gong, S. Palmer, J. Gallacher, T. Marsden, and D. Fone, "A systematic review of the relationship between objective measurements of the urban environment and psychological distress," *Environment International*, vol. 96, pp. 48–57, 2016.

[18] L. Rafaely, S. Carmel, and Y. G. Bachner, "Subjective well-being of visually impaired older adults living in the community," *Aging & Mental Health*, vol. 22, no. 9, pp. 1229–1236, 2018.

[19] Y. Tao, S. Lau, Z. Gou, J. Fu, B. Jiang, and X. Chen, "Privacy and well-being in aged care facilities with a crowded living environment: case study of Hong Kong care and attention homes," *International Journal of Environmental Research and Public Health*, vol. 15, no. 10, p. 2157, 2018.

[20] P. Pollock, M. Stowell-Smith, and M. Göpfert, *Cognitive Analytic Therapy for Offenders: A New Approach to Forensic Psychotherapy*, Routledge, England, UK, 2006.

[21] L. S. Weinhardt, P. Stevens, H. Xie et al., "Transgender and gender nonconforming youths' public facilities use and psychological well-being: a mixed-method study," *Transgender health*, vol. 2, no. 1, pp. 140–150, 2017.

[22] W. Pavot and E. Diener, "The affective and cognitive context of self-reported measures of subjective well-being," *Social Indicators Research*, vol. 28, no. 1, pp. 1–20, 1993.

[23] B. Headey and A. Wearing, "Personality, life events, and subjective well-being: toward a dynamic equilibrium model," *Journal of Personality and Social Psychology*, vol. 57, no. 4, pp. 731–739, 1989.

[24] W. L. Neuman, *Social Research Methods: Qualitative and Quantitative Approaches*, Pearson/Allyn & Bacon, Boston, MA, USA, 2011.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 5229348, 17 pages
https://doi.org/10.1155/2022/5229348

Research Article

Study of the Flood Frequency Based on Normal Transformation in Arid Inland Region: A Case Study of Manas River in North-Western China

Changlu Qiao,[1,2] **Guotao Cai**,[1,3] **Yanxue Liu**,[1,2] **Junfeng Li**,[1,2] **and Fulong Chen**[1,2]

[1]*College of Water Conservancy & Architectural Engineering, Shihezi University, Shihezi 832003, Xinjiang, China*
[2]*Key Laboratory of Modern Water-Saving Irrigation of Xinjiang Production & Construction Group, Shihezi 832000, Xinjiang, China*
[3]*Xinjiang Hami Pumped Storage Power Company Limited, Hami 839000, Xinjiang, China*

Correspondence should be addressed to Guotao Cai; caiguotao@stu.shzu.edu.cn and Yanxue Liu; liuyanxue@stu.shzu.edu.cn

Received 20 May 2022; Revised 19 June 2022; Accepted 22 June 2022; Published 13 July 2022

Academic Editor: Xingsi Xue

Flood disaster is one of the natural disasters which cause the most serious economic losses, the most casualties, and the greatest social impact. Flood frequency analysis is very important for reducing flood disaster. In this paper, based on the flood data of Manas River and tools of Box–Cox and Johnson normal transformation, the nonparametric statistical method for flood frequency analysis is studied in order to analyze the adaptability between it and the rivers in arid region of north-western China. The calculation result of the fitness index is divided into two parts: high flood discharge and low flood discharge. One of the two evaluation indexes has an advantage in fitting, and the number of advantages of the three methods in each part has been counted. After analysis, for the flood peak discharge frequency of rivers in arid region of north-western China, the frequency curve of Johnson transformation fits best with empirical data. The high flood discharge advantage is 6, and the low flood discharge is 4. For the flood volume frequency of rivers in arid region of north-western China, Box–Cox transform fits well with empirical data at the high flood discharge frequency curve, and its advantage is 12; Johnson transformation has a better fit between the low flood discharge frequency curve and empirical data, and its advantage is 12. Therefore, it is the way of improving the precision of flood frequency analysis to use the method of P-III distribution and normal transformation comprehensively.

1. Introduction

In recent years, due to the impact of global climate change and human activities, hydrologic extreme disasters had increased greatly in most countries and regions, resulting in economic, ecological, and even life and property heavy losses [1]. As a hydrologic extreme event, flood disaster is one of the natural disasters which cause the most serious economic losses, the most casualties, and the greatest social impact [2]. As one of the main methods to accurately estimate the design value of hydrologic variables, flood frequency analysis is very important for reducing flood disaster. At present, flood frequency analysis has been widely used in the field of hydrologic design [3]. The current flood frequency analysis methods in China are mainly divided into two categories, the parametric statistical method and the nonparametric statistical method. The parameter statistical method is based on the presumed flood frequency distribution form, and the parameters of the population distribution are obtained through the samples, and the design value under the specified frequency is obtained from the population distribution [4]. In the case of reasonable distribution form assumption, the parameter statistical method can obtain more information from the samples and the calculation results are more accurate. However, if it is unreasonable, the calculation accuracy will be reduced. The nonparametric statistical method avoids the error caused by the assumed distribution form and directly calculates the design value of the specified frequency according to the measured flood samples. The nonparametric statistical method is more flexible and robust than the parametric

statistical method. It is also a research hotspot at present and provides another research way for hydrologic frequency analysis [5–10]. The normal transformation transforms the original skewed distribution of the sample into the normal distribution and then uses its inverse transformation to calculate the design value of hydrologic variables (normal quantile) under the specified frequency. This process does not involve the content of parameter calculation method and takes normal distribution as the intermediate medium. It belongs to the nonparametric statistical method in theory [11–13]. Relevant researches have shown that, after the normal transformation of single variable, the original skewed distribution to the normal distribution is a one-to-one monotonic increasing relationship, and the serry obtained by using the normal transformation can retain the sample information of the original skew distribution more completely [14, 15]. The research of Chen and Song [16] also pointed out that there is a good fitting effect between the design value obtained by the normal transformation and the measured serry. So, the normal transformation can be used in hydrologic frequency analysis.

The application of the normal transformation in the calculation of flood frequency in arid region is less. Johnson transformation is mostly used in quality management statistics [17] and processing nonnormal statistics problems [18]. Box–Cox transformation is often used to improve the skewness and heteroscedasticity of linear regression, and it is more suitable for hydrology than Johnson transformation, such as the research of Liang and Dai [11] and Li et al. [19] and others. As a typical river in arid region of north-western China, hydrologic extreme events often occur at Manasi River. In this paper, based on the flood data of Manas River and tools of Box–Cox and Johnson normal transformation, the nonparametric statistical method for flood frequency analysis is studied in order to analyze the adaptability between it and the rivers in arid region of north-western China.

2. Materials and Methods

2.1. Box–Cox Transformation. Box–Cox transformation is a normal transformation model proposed by Box and Cox in 1964. The model is [20–25]

$$Y = \begin{cases} \dfrac{X^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \log X, & \lambda = 0. \end{cases} \tag{1}$$

The inverse transformation of the model is

$$X = \begin{cases} (1 + \lambda Y)^{1/\lambda}, & \lambda \neq 0, \\ e^{Y}, & \lambda = 0, \end{cases} \tag{2}$$

where $X = \{x_1, x_2, \ldots x_n\}$ is the original serry to be transformed; $Y = \{y_1, y_2, \ldots, y_n\}$ is the output serry after transformation; λ is the transformation parameter, $\lambda \epsilon [-5, +5]$. The method to determine the best λ is when $\lambda \epsilon [-5, +5]$, the optimal value of λ is the one where the minimum standard deviation of the Z serry defined by equation (3) is obtained [26]. When $\lambda = 0$, the transformation is logarithmic transformation, $\lambda = -1$ is the reciprocal transformation, and $\lambda = 0.5$ is the square root transformation.

$$Z = \begin{cases} \dfrac{X^{\lambda} - 1}{\lambda g^{\lambda - 1}}, & \lambda \neq 0, \\ g \cdot Ln(X), & \lambda = 0, \end{cases} \tag{3}$$

where g is the geometric mean of the original serry, X is the original data serry to be transformed, and Z is the output serry after transformation.

Box–Cox normal transformation requires that each item of the serry to be transformed is greater than 0. That is $x_i > 0$. Each item of hydrologic serry is greater than 0, so it meets the transformation requirements.

2.2. Johnson Transformation. Johnson transformation is a normal transformation model based on three distribution curves proposed by Johnson in 1949. The model is shown in Table 1 [27–30].

Chou et al. [32–34]. $X = \{x_1, x_2, \ldots x_n\}$ is the original serry to be transformed; $Y = \{y_1, y_2, \ldots y_n\}$ is the output serry after transformation; ε and γ are position control parameters; λ and η are scale parameters and are generally positive. In the Johnson normal transformation, the calculation of the parameters to be estimated is based on the method proposed by Hill et al. [31] and Chou et al. [32–34].

2.3. Study Area. The Manas River is located in the northern foot of the middle section of Tianshan and on the southern edge of the Junggar Basin. It is the largest river on the northern slope of the Tianshan. It originates from the Erenhabirga Mountains on the northern slope of the Tianshan. It is about 324 km in length, and the drainage area is about 5156 km^2 [35]. Kenswat Hydrologic Station is midstream of Manas River, which was built in 1955. The station controls the flow of Manas River (Figure 1). The hydrologic data have been compiled and reviewed by the Hydrology and Water Resources Bureau with reliable accuracy [36].

2.4. Data Acquisition and Processing. In this paper, the measured flood data of Kenswat Hydrologic Station with the longest observation time of Manas River are used as the measured serry. Kenswat reservoir is one kilometer upstream of Kenswat Hydrological Station. The construction of the reservoir started on August 7, 2009, and officially began to impound on December 6, 2014. Since the impoundment of the reservoir, the consistency of hydrologic data of Kenswat Hydrologic Station had been destroyed. Therefore, the data period selected in this paper is from 1955 to 2014.

According to the data analysis of the station, the average annual flood peak discharge of the Manas River is 356 m^3/s. The measured maximum flood peak discharge is 1095 m^3/s (August 2, 1999), the second is 758 m^3/s (July 28, 1966), and the third is 735 m^3/s (July 18, 1996). The annual maximum sampling method is used to select flood peak discharge

Table 1: Johnson transformation model.

Type	Transformation model	Inverse transformation model	Parameter constraints	X constraints
SB	$Y = \gamma + \eta \ln(X - \varepsilon/\lambda + \varepsilon - X)$	$X = \varepsilon + (\lambda + \varepsilon)e^{Y+\gamma/\eta}/1 + e^{Y+\gamma/\eta}$	$\eta, \lambda > 0$ $-\infty < \gamma < +\infty$ $-\infty < \varepsilon < +\infty$	$\varepsilon < X < \varepsilon + \lambda$
SL	$Y = \gamma + \eta \ln(X - \varepsilon)$	$X = e^{Y-\gamma/\eta} + \varepsilon$	$\eta > 0$ $-\infty < \gamma < +\infty$ $-\infty < \varepsilon < +\infty$	$X > \varepsilon$
SU	$Y = \gamma + \eta \text{arc } sinh(X - \varepsilon/\lambda)$	$X = \lambda \sinh(Y - \gamma/\eta) + \varepsilon$	$\eta, \lambda > 0$ $-\infty < \gamma < +\infty$ $-\infty < \varepsilon < +\infty$	$-\infty < X < +\infty$
Note		$\text{arc } sin\, hx = \ln[x + (x^2 + 1)^{0.5}]$		

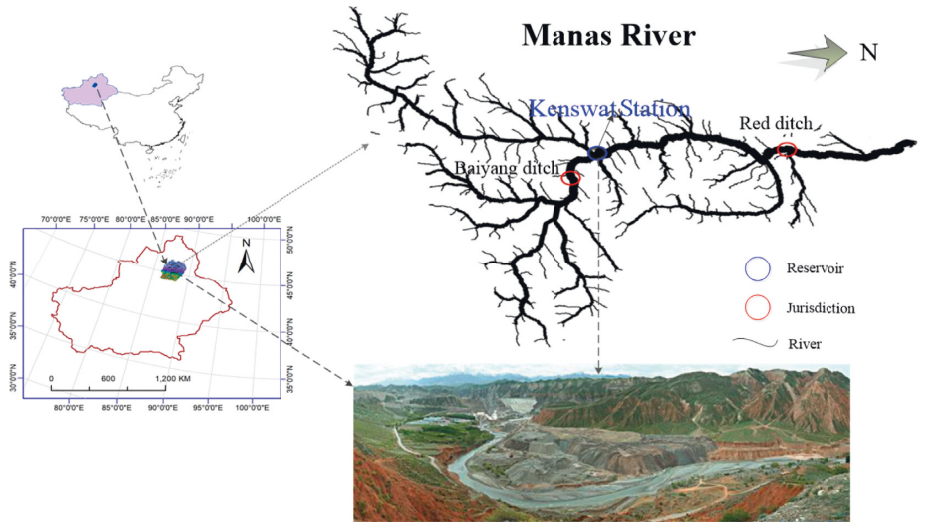


Figure 1: Locations of Manas river and Ken Swat station.

samples, the unified sample method is adopted, and the Weibull equation is used to determine the empirical frequency. Because the Manas River is a small river, when selecting flood volume samples, 1 d, 3 d, and 5 d are used as the flood volume calculation period.

3. Results and Discussions

3.1. Detection and Correction of the Mutation

3.1.1. Detection of the Mutation

(1) Flood Peak Discharge Serry. In order to ensure the accuracy of detecting the mutation, three methods are selected to detect the mutation at the same time, and the mutation point is finally determined by comprehensive analysis.

According to calculation and analysis, the skipping mutation point of Lee–Heghinan test is 1995, and the skipping mutation point of ordered clustering test and sliding T test is 1993. Combining other literature and test results, the flood peak discharge serry mutation point is determined to be 1993, and so the serry is divided into two subseries from the mutation point. The subseries before the mutation point shows a downward trend, and its average value is 334.08 m^3/s. The subseries after the mutation point also shows a downward trend, and its average value is 442.87 m^3/s. The average value of the two subseries is quite different, showing skipping change. According to calculation and analysis, the main mutation type of flood peak discharge serry is identified as skipping mutation, and the results of detecting the mutation are shown in Figures 2(a) and 2(b).

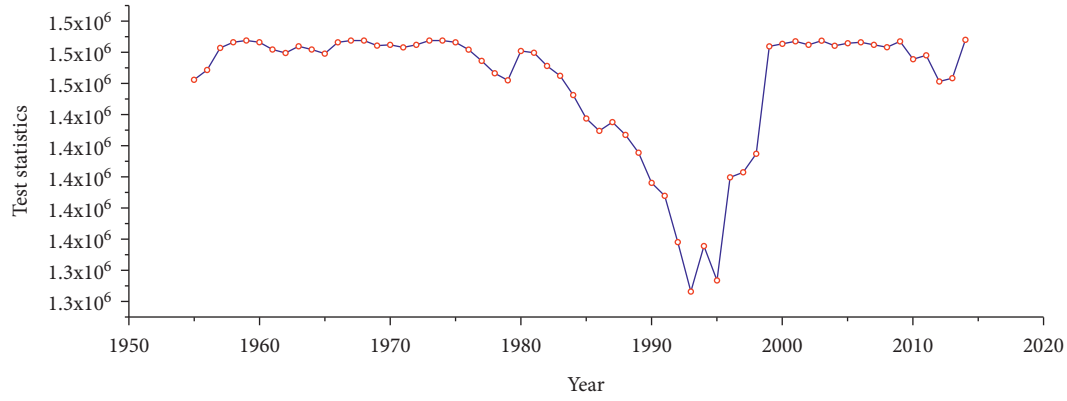
(2) Flood Volume Serry. The method of detecting the mutation of the flood volume serry is the same as that of the flood peak discharge. The test results are shown in Figures 2(c)–2(h).

It can be seen from Figures 2(c)–2(h) that the maximum 1 d, 3 d, and 5 d flood volume serry mutation points are all in 1993. It is consistent with the test results of flood peak serry. So, it can be determined that the main mutation type of Manas River flood serry is skipping mutation and the mutation point is in 1993.

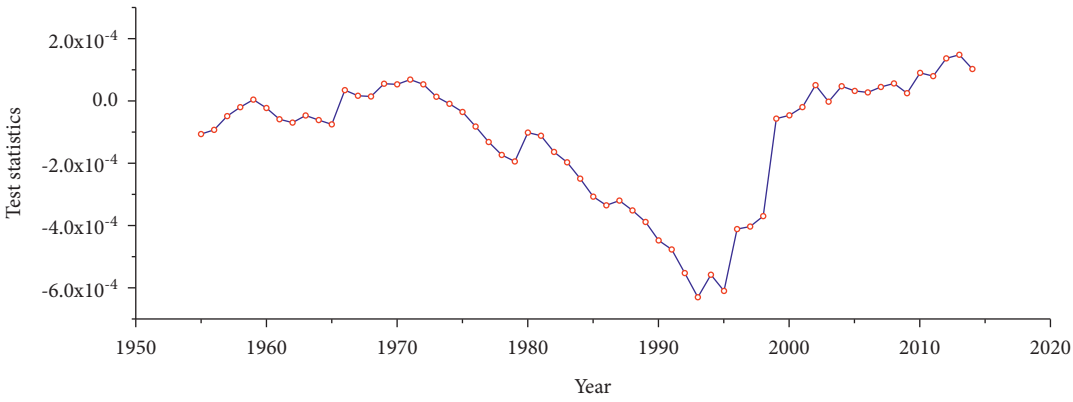
3.1.2. Correction of the Mutation

(1) Flood Peak Discharge Serry. In this paper, the decomposition synthesis theory proposed by Xie Ping is adopted as the method of correcting the skipping mutation, and the calculation process is based on the previous researches [37–39].

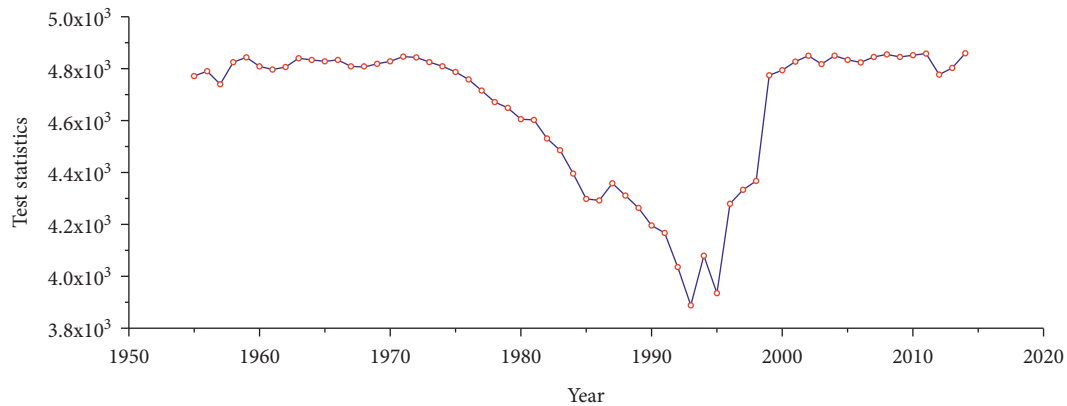
According to the decomposition synthesis theory proposed by Xie Ping, the equation of correcting the skipping mutation is



(a)

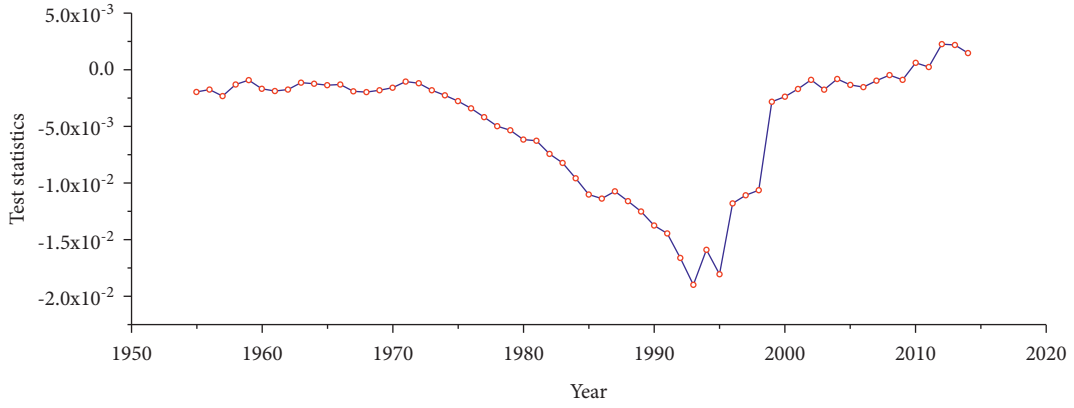


(b)

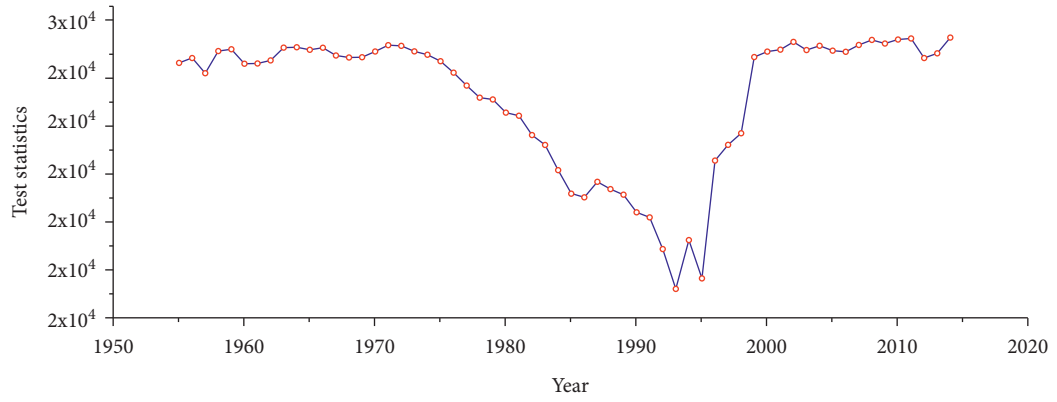


(c)

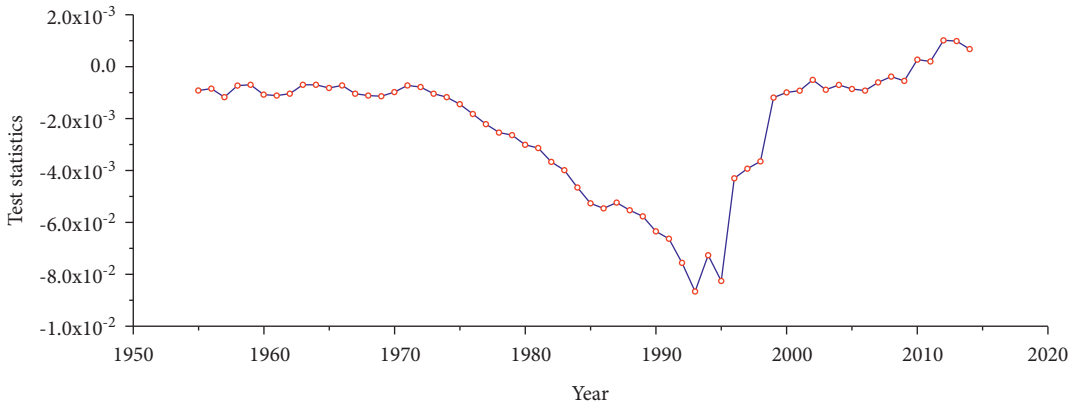
Figure 2: Continued.



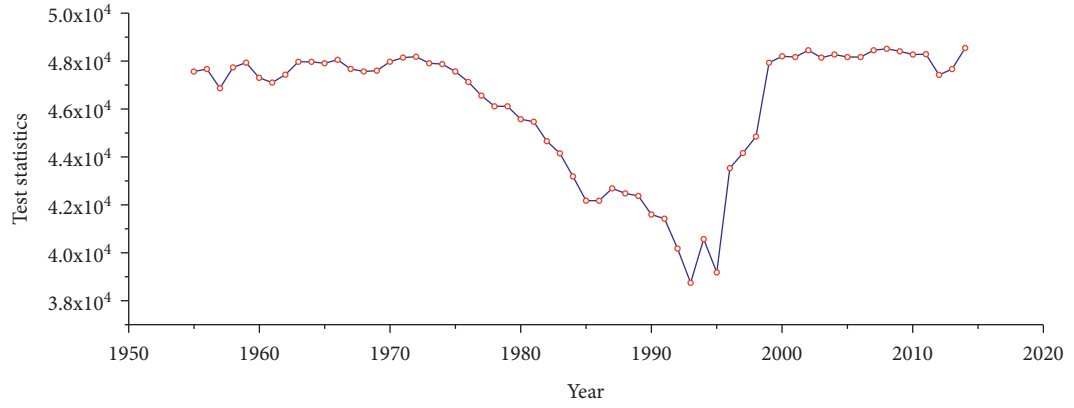
(d)



(e)

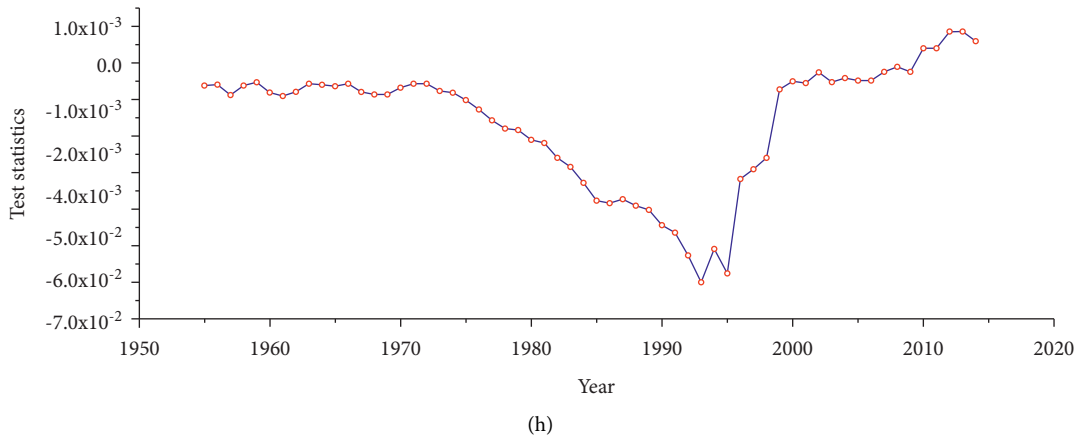


(f)



(g)

Figure 2: Continued.



(h)

Figure 2: Results of detecting the mutation of the flood serry. (a) and (b) The results of ordered cluster test and sliding T test of the annual maximum flood peak discharge serry, (c) and (d) the results of the ordered clustering test and the sliding T test of the maximum 1 d flood volume, (e) and (f) the results of ordered clustering test and sliding T test with the maximum 3 d volume, and (g) and (h) the results of ordered clustering test and sliding T test with the maximum 5 d volume.

$$x_k' = x_k - 108.79, \quad (k = j+1, j+2, \ldots, n), \tag{4}$$

where x_k' is the item of the serry after correcting the mutation; x_k is the item of the serry before correcting the mutation; and j is the year of mutation (here is 1993).

Based on the subseries before the mutation point, the subseries after the mutation point is corrected by equation (4). The corrected synthetic serry is detected again, and no mutation point is found, so the corrected serry meets the premise of consistency. The corrected flood peak discharge serry is shown in Figure 3(a).

(2) Flood Volume Serry. The method of correcting the mutation of the flood volume serry is the same as that of the flood peak discharge. The corrected results are shown in Figures 3(b)–3(d).

3.2. Normal Detection and Transformation

3.2.1. Normal Detection

(1) Flood Peak Discharge Serry. The normal detection is carried out on the flood peak discharge serry after correcting the mutation. The detection methods are the nonparametric Shapiro–Wilk test (W test) and normal P–P diagram method. The normal P–P test results are shown in Figure 4(a), and the W test results are shown in Table 2.

According to Figure 4(a) and Table 2, the normality of the flood peak discharge serry is not significant, so the normal transformation cannot be directly used to calculate the flood peak discharge frequency. Therefore, the skewed serry is transformed into the normal serry by the methods of Box–Cox transformation and Johnson transformation, and the normal test results of the transformed flood peak discharge serry are shown in Figures 5(a) and 5(b).

It can be seen from Figures 5(a) and 5(b) and the results of normal transformation for the corrected flood peak discharge serry that the effect of the Box–Cox normal transformation is poor, and the effect of Johnson normal transformation is better.

(2) Flood Volume Serry. For the normal test of the flood volume serry, the method is the same as that of the flood peak discharge.

(I) Maximum 1 d flood volume serry

The normal test is performed on the maximum 1 d flood volume after correcting the mutation, and the test results are shown in Figure 4(b) and Table 2. The normal test results of the maximum 1 d flood volume serry after normal transformation are shown in Figure 5(c) and 5(d).

(II) Maximum 3 d flood volume serry

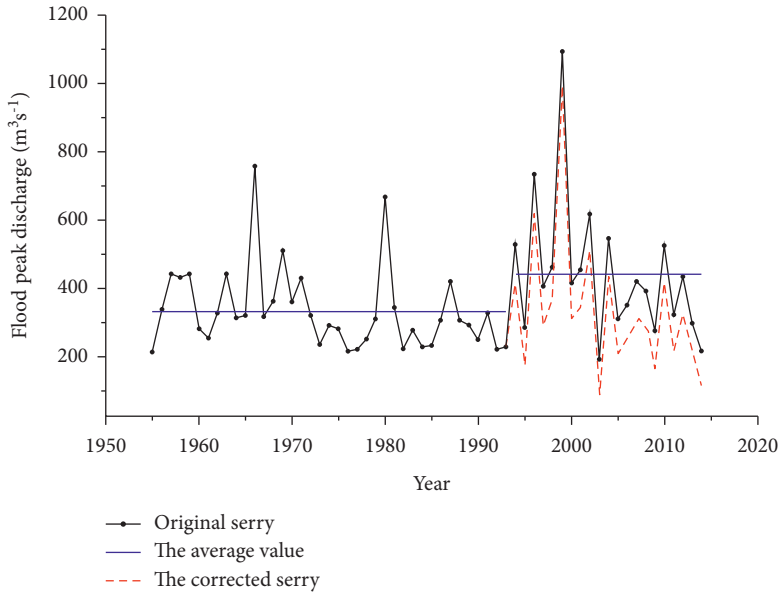
The normal test is performed on the maximum 3 d flood volume after correcting the mutation, and the test results are shown in Figure 4(c) and Table 2. The normal test results of the maximum 3 d flood volume serry after normal transformation are shown in Figure 5(e) and 5(f).

(III) Maximum 5 d flood volume serry

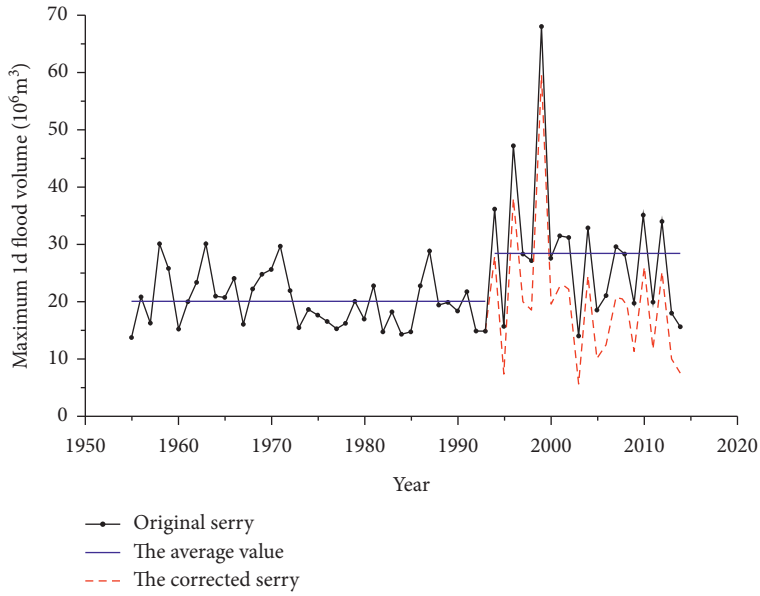
The normal test is performed on the maximum 5 d flood volume after correcting the mutation, and the test results are shown in Figure 4(d) and Table 2. The normal test results of the maximum 5 d flood volume serry after normal transformation are shown in Figures 5(g) and 5(h).

3.2.2. Normal Transformation

(1) Flood Peak Discharge Serry. After Box–Cox normal transformation for the corrected flood peak discharge serry, the new serry passed the normal test. The P-value of 95% confidence interval of W test is equal to 0.06. The mean value of the new serry is 13.78, and the standard deviation is 1.998.



(a)



(b)

Figure 3: Continued.

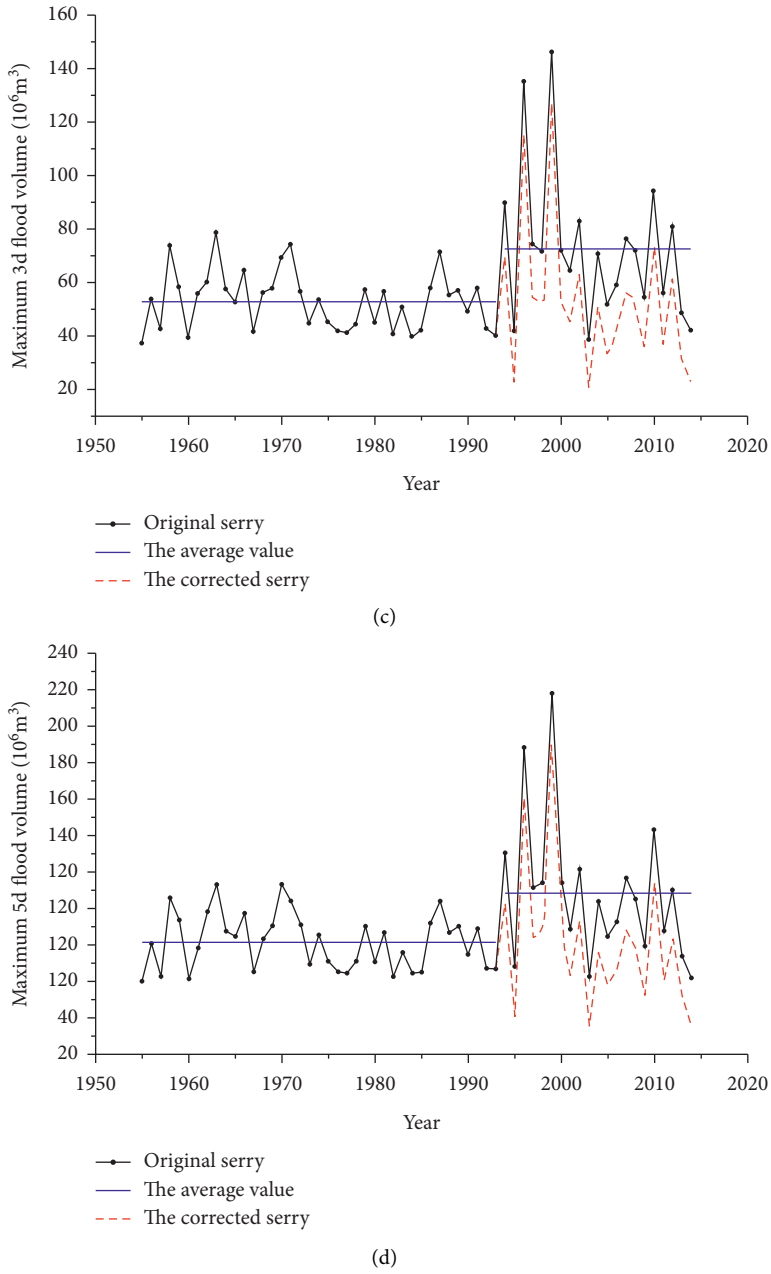


Figure 3: Results of correcting the mutation of flood serry. (a) The correction result of the annual maximum flood peak discharge mutation, (b) the correction result of the maximum 1 d flood volume mutation, (c) the correction result of the maximum 3 d flood volume mutation, and (d) the correction result of the maximum 5 d flood volume mutation.

By comparing S_B, S_L, and S_U, the three types of Johnson normal transformation for the corrected flood peak discharge serry, we found that S_U transformation is the best. After S_U type of Johnson normal transformation for the corrected flood peak discharge serry, the new serry passed the normal test. The P-value of 95% confidence interval of W test is equal to 0.87. The mean value of the new serry is 0.06259, and the standard deviation is 0.9161. The parameters of the S_U-type Johnson normal transformation for the corrected flood peak serry are as follows: $\gamma = 0.831159$, $\eta = 1.12611$, $\varepsilon = 221.007$, $\lambda = 92.8704$. So, the optimal equation of S_U-type Johnson normal transformation is

$$Y = -0.831159 + 1.12611 \operatorname{arc}\sinh\left(\frac{X - 221.007}{92.8704}\right). \quad (5)$$

The best inverse normal transformation is

$$X = 92.8704 \sinh\frac{Y + 0.831159}{1.12611} + 221.007. \quad (6)$$

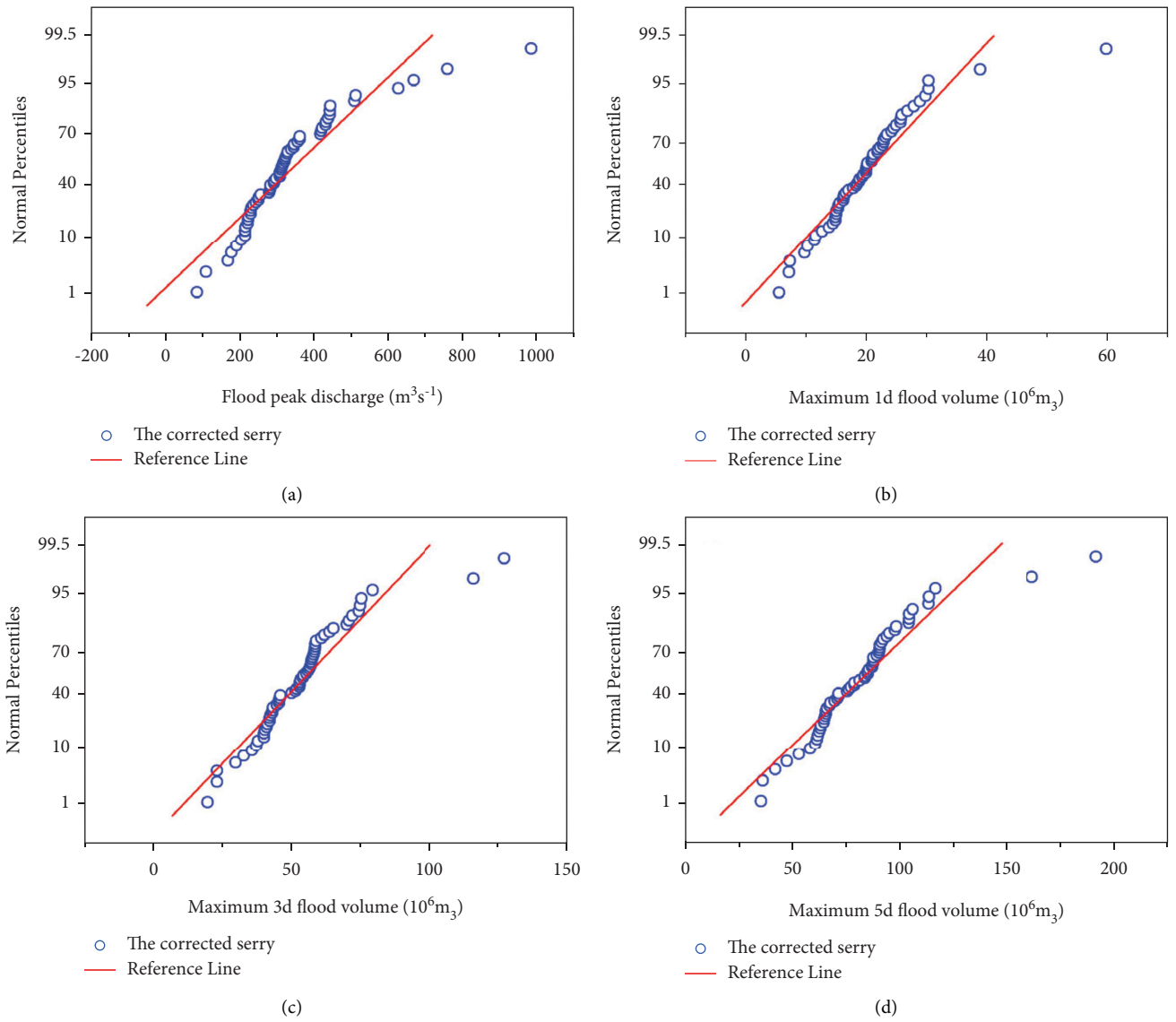


Figure 4: Normal P–P test results of the flood serry. (a) The normal test chart of the annual maximum flood peak discharge after correcting the mutation, (b) the normal test chart of the maximum 1 d flood volume after correcting the mutation, (c) the normal test chart of the maximum 3 d flood volume after correcting the mutation, and (d) the normal test chart of the maximum 5 d flood volume after correcting the mutation.

Table 2: W test results of flood.

Test variable	DF	Statistic	P-value	Decision at level (5%)
Flood peak discharge	60	0.84	1.76×10^{-6}	Reject normality
Maximum 1 d flood volume	60	0.86	6.17×10^{-6}	Reject normality
Maximum 3 d flood volume	60	0.87	1.58×10^{-5}	Reject normality
Maximum 5 d flood volume	60	0.88	2.2×10^{-5}	Reject normality

(2) Flood Volume Serry

(I) Maximum 1 d flood volume serry

The normal transformation results of maximum 1 d flood volume serry are that the best transformation parameter of λ is equal to 0 for Box–Cox normal transformation and the best normal transformation equations are as follows for Johnson normal transformation.

$$Y = -0.383021 + 1.5283\text{arc}\,sinh\left(\frac{X - 16.8127}{8.74075}\right),$$
$$X = 8.740775\ \sinh\frac{Y + 0.383021}{1.5283} + 16.8127. \tag{7}$$

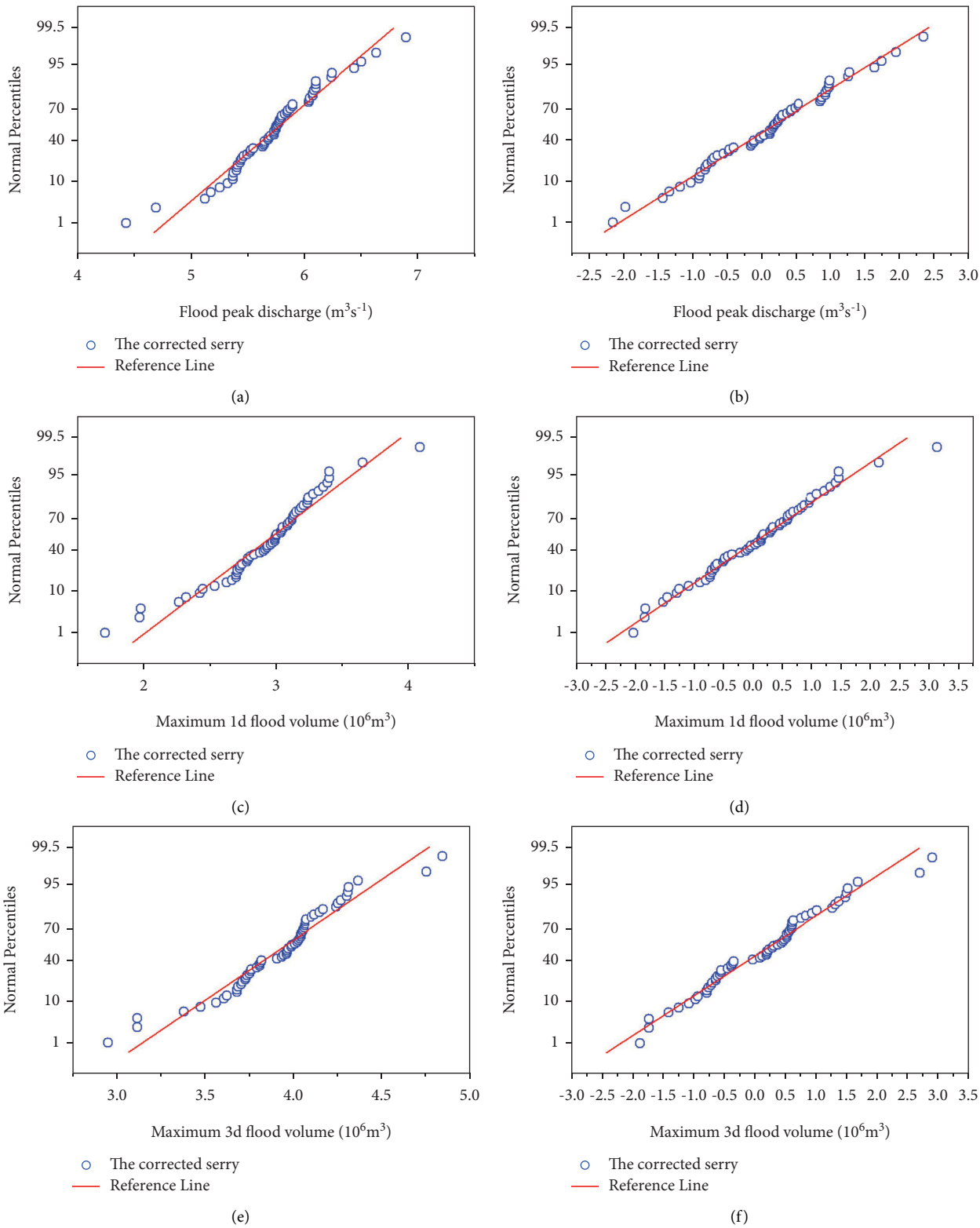


Figure 5: Continued.

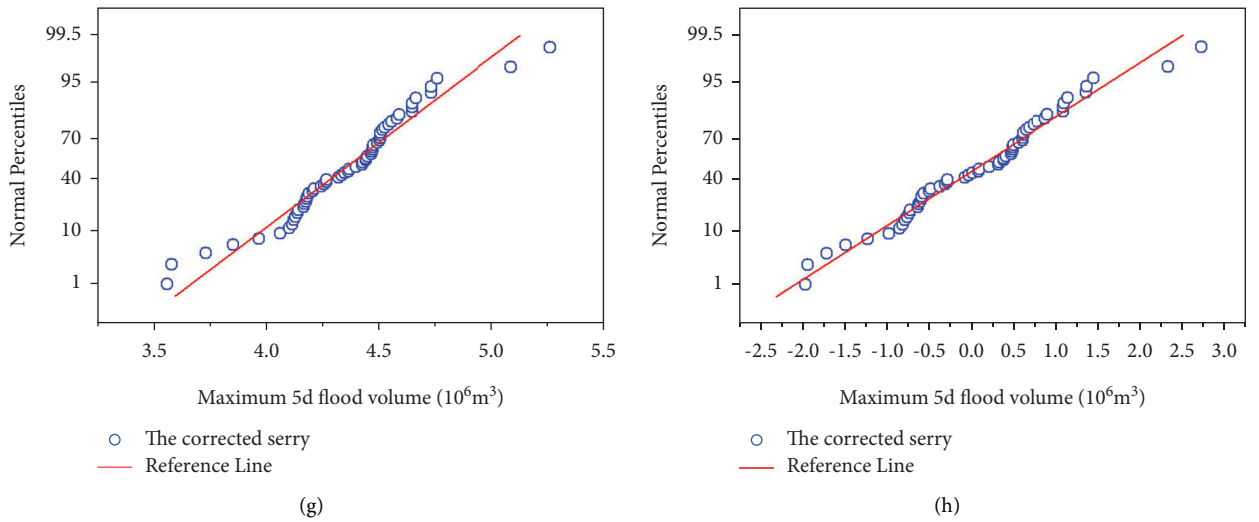


Figure 5: The normal test results of the transformed flood serry. (a) and (b) The Box–Cox transformation and Johnson transformation of the annual maximum flood peak discharge, (c) and (d) the Box–Cox transformation and Johnson transformation of the maximum 1 d flood volume, (e) and (f) the Box–Cox transform and Johnson transformation of the maximum 3 d flood volume, and (g) and (h) the Box–Cox transform and Johnson transformation of the maximum 5 d flood volume.

(II) Maximum 3 d flood volume serry

The normal transformation results of maximum 3 d flood volume serry are that the best transformation parameter of λ is equal to 0 for Box–Cox normal transformation and the best normal transformation equations are as follows for Johnson normal transformation.

$$Y = -0.118591 + 1.36577\text{arc}\,sinh\left(\frac{X - 48.5286}{17.466}\right),$$
$$X = 17.466\ \sinh\frac{Y + 0.118591}{1.36577} + 48.5286. \tag{8}$$

(III) Maximum 5 d flood volume serry

The normal transformation results of maximum 5 d flood volume serry are that the best transformation parameter of λ is equal to 0 for Box–Cox normal transformation and the best normal transformation equations are as follows for Johnson normal transformation.

$$Y = -0.54766 + 1.48425\text{arc}\,sinh\left(\frac{X - 66.3067}{28.1313}\right),$$
$$X = 28.1313\ \sinh\left(\frac{Y + 0.54766}{1.48425}\right) + 66.3067. \tag{9}$$

3.3. Flood Frequency

3.3.1. Flood Peak Discharge Frequency Curve. Using the inverse transformation model of Box–Cox and Johnson normal transformation given above, the design value (normal quantile) of normal distribution under specified frequency is used to deduce the design flood peak discharge corresponding to each frequency of the original distribution and draw the flood peak discharge frequency curve. At the same time, based on P-III distribution, the design flood peak discharge under each frequency is calculated by using the Optimization Curve-Fitting Method [40, 41], and the P-III distribution of flood peak discharge frequency curve can be drawn. The design flood peak discharge under specified frequency of normal distribution obtained by two normal transformations is shown in Tables 3 and 4. The three kinds of flood peak discharge frequency curves are shown in Figure 6(a).

It can be seen from Figure 6(a) that the flood peak discharge frequency curve deduced by Johnson normal transformation fits the measured value best. The fitting goodness of Box–Cox transformation and P-III distribution cannot be seen directly, so further quantitative calculation is needed.

3.3.2. Flood Volume Frequency Curve. Based on Box–Cox normal transformation, Johnson normal transformation, and P-III distribution, the flood volume frequency curves of maximum 1 d, 3 d, and 5 d are deduced. And the results are shown in Figures 6(b)–6(d).

3.4. Goodness-of-Fit Calculation. Using the two evaluation indexes of Mean Square Error (MSE) and Residual Sum of Squares (RSS), in this paper, the goodness of fitness between the measured values of high flood discharge (10%, 30%, and 50%) and low flood discharge (10% and 30%) with the corresponding designed values of different methods are calculated [3, 42, 43]. The results are shown in Table 5.

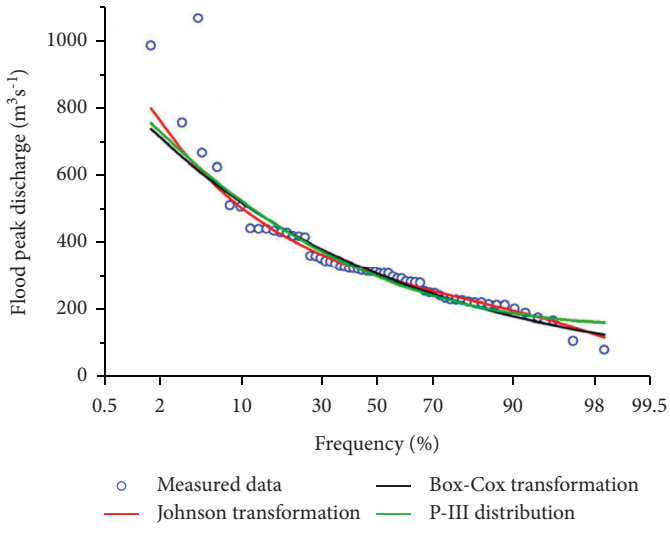
According to the calculation results of RSS evaluation index of flood peak frequency, the order of goodness of fitness between the measured value and the designed value is Johnson transformation > Box–Cox transformation > P-III distribution. And the calculation result of MSE evaluation index is the same as the RSS evaluation index.

Table 3: The design flood peak discharge at specified frequency of normal distribution (Box–Cox).

Box–Cox transformation							
1	4.8478894	16	5.4650952	31	5.7356066	46	6.0101125
2	4.9688762	17	5.4852734	32	5.7521198	47	6.0319379
3	5.0462946	18	5.5050398	33	5.7694154	48	6.0547928
4	5.1051820	19	5.5244768	34	5.7862992	49	6.0786772
5	5.1537744	20	5.5433372	35	5.8035948	50	6.1033852
6	5.1953662	21	5.5618682	36	5.8208904	51	6.1301522
7	5.2324282	22	5.5799874	37	5.8381860	52	6.1581546
8	5.2653722	23	5.5981066	38	5.8558934	53	6.1886278
9	5.2958454	24	5.6158140	39	5.8740126	54	6.2215718
10	5.3238478	25	5.6331096	40	5.8921318	55	6.2586338
11	5.3506148	26	5.6504052	41	5.9106628	56	6.3002256
12	5.3753228	27	5.6677008	42	5.9295232	57	6.3488180
13	5.3992072	28	5.6845846	43	5.9489602	58	6.4077054
14	5.4220621	29	5.7018802	44	5.9687266	59	6.4851238
15	5.4438875	30	5.7183934	45	5.9889048	60	6.6061106

Table 4: The design flood peak discharge at specified frequency of normal distribution (Johnson).

Johnson transformation							
1	−1.8931003	16	−0.5200496	31	0.0817365	46	0.6924088
2	−1.6239501	17	−0.4751607	32	0.1184721	47	0.7409621
3	−1.4517233	18	−0.4311879	33	0.1569483	48	0.7918056
4	−1.3207210	19	−0.3879480	34	0.1945084	49	0.8449394
5	−1.2126212	20	−0.3459906	35	0.2329846	50	0.8999054
6	−1.1200951	21	−0.3047661	36	0.2714608	51	0.9594519
7	−1.0376461	22	−0.2644577	37	0.3099370	52	1.0217467
8	−0.9643581	23	−0.2241493	38	0.3493293	53	1.0895381
9	−0.8965667	24	−0.1847570	39	0.3896377	54	1.1628261
10	−0.8342719	25	−0.1462808	40	0.4299461	55	1.2452751
11	−0.7747254	26	−0.1078046	41	0.4711706	56	1.3378012
12	−0.7197594	27	−0.0693284	42	0.5131280	57	1.4459010
13	−0.6666256	28	−0.0317683	43	0.5563679	58	1.5769033
14	−0.6157821	29	0.0067079	44	0.6003407	59	1.7491301
15	−0.5672288	30	0.0434435	45	0.6452296	60	2.0182803



(a)

Figure 6: Continued.

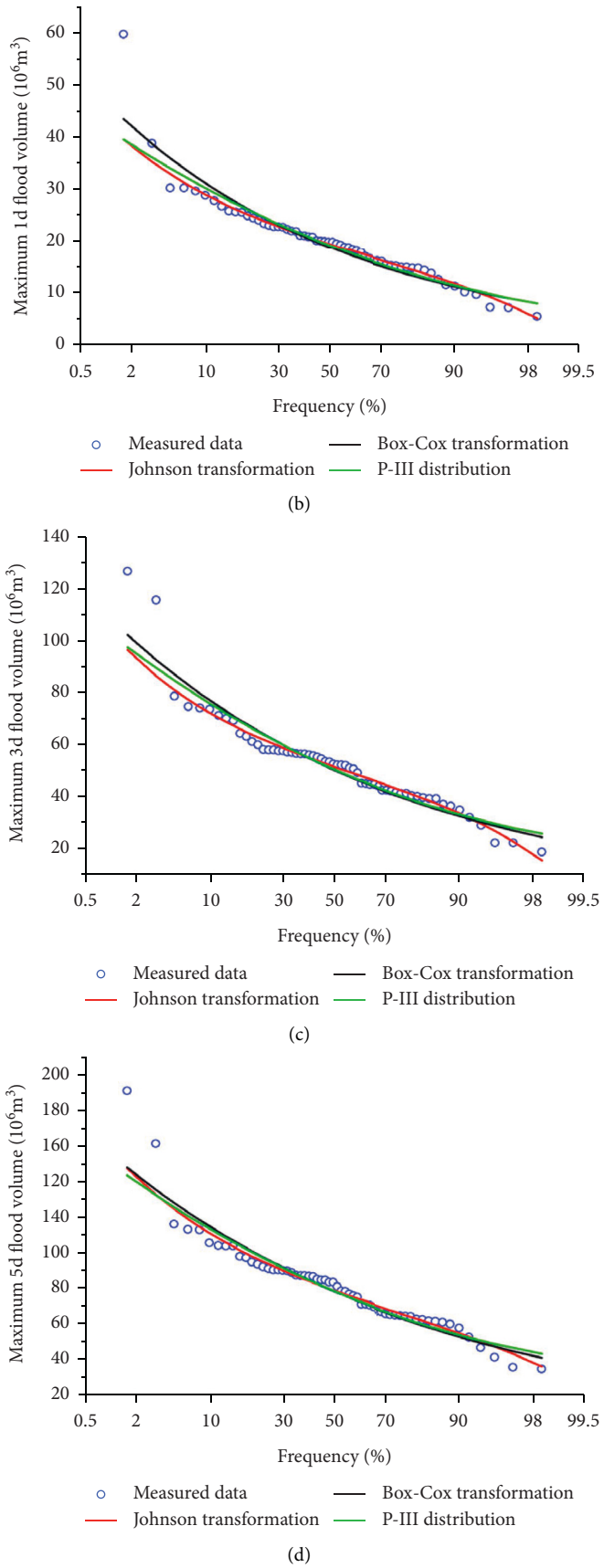


Figure 6: The three flood frequency curves. (a) The calculation result of the annual maximum flood peak discharge frequency, (b) the calculation result of the maximum 1 d flood volume frequency, (c) the calculation result of the maximum 3 d flood volume frequency, and (d) the calculation result of the maximum 5 d flood volume frequency.

Table 5: The calculation results of goodness of fitness.

Hydrologic serry			RSS			MSE		
			P-III distribution	Box–Cox transformation	Johnson transformation	P-III distribution	Box–Cox transformation	Johnson transformation
Flood peak	High	10%	67034.13	79364.90	48848.57	11172.35	13227.48	8141.43
		30%	77033.56	89651.90	53797.42	4279.64	4980.66	2988.75
		50%	79019.54	93676.98	54608.96	2633.98	3122.57	1820.30
	Low	10%	10112.16	4281.46	3005.88	1685.36	713.58	500.98
		30%	11622.91	6278.01	3621.84	645.72	348.78	201.21
Maximum 1 d flood volume	High	10%	437.10	323.04	427.55	73.00	53.84	71.26
		30%	458.85	353.17	430.11	25.49	19.62	23.90
		50%	459.82	355.81	430.50	15.33	11.86	14.35
	Low	10%	17.57	16.75	5.90	2.93	2.79	0.98
		30%	26.81	33.33	8.39	1.49	1.85	0.47
Maximum 3 d flood volume	High	10%	1619.24	1307.61	1771.43	269.87	217.93	295.24
		30%	1806.61	1532.90	1834.97	100.37	85.16	101.94
		50%	1832.43	1564.00	1846.39	61.08	52.13	61.55
	Low	10%	142.75	103.33	33.81	23.79	17.22	5.63
		30%	186.77	172.85	51.87	10.38	9.60	2.88
Maximum 5 d flood volume	High	10%	3350.42	2891.58	2884.61	558.40	481.93	480.77
		30%	3604.41	3220.96	2964.18	200.25	178.94	164.68
		50%	3689.12	3308.46	3055.62	122.97	110.28	101.85
	Low	10%	279.59	191.31	120.71	46.60	31.89	20.12
		30%	360.82	313.14	167.53	20.05	17.40	9.31

Table 6: Statistical tables of goodness of fitness.

Variables	Method	High	Low
Flood peak	P-III	0	0
	Box–Cox	0	0
	Johnson	6	4
Maximum 1 d flood volume	P-III	0	0
	Box–Cox	6	0
	Johnson	0	4
Maximum 3 d flood volume	P-III	0	0
	Box–Cox	6	0
	Johnson	0	4
Maximum 5 d flood volume	P-III	0	0
	Box–Cox	0	0
	Johnson	6	4

According to the calculation results of RSS evaluation index of flood volume frequency, for the maximum 1 d flood volume serry, the order of goodness of fitness between the measured value and the designed value is Box–Cox transformation > Johnson transformation > P-III distribution at high flood discharge (10%, 30%, and 50%), and Johnson transformation > Box–Cox transformation > P-III distribution at low flood discharge (10% and 30%), and the calculation results of MSE evaluation index are the same as the RSS evaluation index.

For the maximum 3 d flood volume serry, the order of goodness of fitness between the measured value and the designed value is Box–Cox transformation > P-III distribution > Johnson transformation at high flood discharge (10%, 30%, and 50%), and Johnson transformation > Box–Cox transformation > P-III distribution at low flood discharge (10% and 30%), and the calculation results of MSE evaluation index are the same as RSS evaluation index.

For the maximum 5 d flood volume serry, the order of goodness of fitness between the measured value and the designed value is Johnson transformation > Box–Cox transformation > P-III distribution both at high flood discharge (10%, 30%, and 50%) and low flood discharge (10% and 30%), and the calculation results of MSE evaluation index are the same as the RSS evaluation index.

The calculation results of goodness of fitness are divided into two parts: high flood discharge and low flood discharge. The number of advantages of the three methods in each part is counted. The results are shown in Table 6.

4. Conclusion

In this paper, 60 years measured flood data of Manas River in north-western China are used to study the adaptability of flood frequency analysis based on normal transformation to Manas River, by comparing with the traditional P-III distribution. The conclusions are as follows:

(1) For the flood peak discharge frequency, Johnson transformation is best at the fitting accuracy between frequency curve with empirical data among the three methods of Johnson transformation, Box–Cox transformation, and P-III distribution. So, Johnson transformation is more excellent and adaptive than the traditional P-III distribution in analysis of the flood peak discharge frequency of Manas River.

(2) For the flood volume frequency, Johnson transformation has strong adaptability to stable hydrologic serry, but Box–Cox transformation has strong adaptability to unstable high flood discharge of hydrologic serry. The main reason is that Johnson transformation is a multiparameter transformation and that Box–Cox transformation is a single-parameter transformation. The multiparameter transformation is more stable than the single-parameter one, so does the inverse transformation. Therefore, there is a higher fitting accuracy between all part of frequency curve with empirical data of Johnson transformation than Box–Cox transformation. But at the high flood discharge part of frequency curve, the fitting accuracy of Box–Cox transformation is higher than that of Johnson transformation.

(3) The P-III distribution frequency analysis method has its unique advantages, and the calculation results are relatively stable. When the distribution of the original hydrologic serry is more consistent with the P-III distribution, more information can be obtained from the sample, and the calculation results are more accurate, but when the distribution of the original hydrologic serry is not consistent with the P-III distribution, it will cause more errors. The results of this paper show that the normal transformation method has certain advantages and rationality at flood frequency analysis of rivers in arid island region of north-western China. Therefore, the precision of flood frequency analysis can be improved by using P-III distribution and normal transformation comprehensively.

Data Availability

The data of Kenswat Hydrologic Station with the longest observation time of Manas River are provided by Kenswat Hydrological Station. The data period selected in this paper is from 1955 to 2014.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was financially supported by the Xinjiang Production and Construction Corps (nos. 2022DB024 and 2021AA003) and is gratefully acknowledged.

References

[1] I. Yu, M. L. A. Felix, L. H. Kim, and S. Jeong, "Analysis on flood frequency and water quality variations induced by abnormal climate," *Desalination and Water Treatment*, vol. 53, no. 9, pp. 2339–2353, 2015.

[2] Q. Schiermeier, "Increased flood risk linked to global warming," *Nature*, vol. 470, no. 7334, 316 pages, 2011.

[3] C. Q. Ye, X. H. Chen, Q. X. Shao, and J. M. Zhang, "A comparison study on distribution for flood frequency analysis focus on the observed flow data in the high flow part (in Chinese with English abstract)," *Journal of Hydraulic Engineering*, vol. 44, pp. 68–76, 2013.

[4] Y. F. Chen, Y. Hou, P. Van Gelder, and S. Zhigui, "Study of parameter estimation methods for Pearson-III distribution in flood frequency analysis," *Extremes of the Extremes: Extraordinary Floods*, vol. 271, pp. 263–269, 2002.

[5] H. Vittal, J. Singh, P. Kumar, and S. Karmakar, "A framework for multivariate data-based at-site flood frequency analysis: essentiality of the conjugal application of parametric and nonparametric approaches," *Journal of Hydrology*, vol. 525, pp. 658–675, 2015.

[6] K. Adamowski, "A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies," *Journal of Hydrology*, vol. 108, pp. 295–308, 1989.

[7] S. L. Guo, "Nonparametric estimation of flood frequencies," *Water Resources and Power*, vol. 9, pp. 324–332, 1991.

[8] S. B. Song, "Research challenges and suggestions of hydrological frequency calculation (in Chinese with English abstract)," *Journal of Water Resources and Architectural Engineering*, vol. 17, pp. 12–18, 2019.

[9] P. Vijay and W. G. Singh, "Strupczewski, on the status of flood frequency analysis," *Hydrological Processes*, vol. 16, pp. 3737–3740, 2002.

[10] K. Adamowski, "Nonparametric kernel estimation of flood frequencies," *Water Resources Research*, vol. 21, no. 11, pp. 1585–1590, 1985.

[11] Z. M. Liang and C. J. Dai, "Comparisons of two normalization approaches in hydrologic analytic calculation," *Hydroelectric Energy*, vol. 23, pp. 1–3+89, 2005.

[12] J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis, "Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form," *Water Resources Research*, vol. 15, no. 5, pp. 1049–1054, 1979.

[13] A. I. Fleishman, "A method for simulating non-normal distributions," *Psychometrika*, vol. 43, no. 4, pp. 521–532, 1978.

[14] X. Y. Chen and Y. K. Tung, "Investigation of polynomial normal transform," *Structural Safety*, vol. 25, no. 4, pp. 423–445, 2003.

[15] W. Xu, "Efficiency of Bayesian probabilistic hydrological forecast system based on Box-Cox transformation (in Chinese with English abstract)," *Journal of Hydroelectric Engineering*, vol. 37, pp. 15–23, 2018.

[16] D. F. Chen and S. B. Song, "Investigation of annual runoff frequency calculation based on polynomial normal transformation (in Chinese with English abstract)," *Journal of Hydroelectric Engineering*, vol. 38, pp. 58–69, 2019.

[17] D. B. Zhuo and G. P. Xia, "Non-normal Processes Quality Control Using the Best Johnson Transformation," in *Proceedings of the 7th international conference on industrial management*, pp. 177–182, Japan, November 2004.

[18] W. M. Luh and J. H. Guo, "Using Johnson's transformation and robust estimators with heteroscedastic test statistics: an examination of the effects of non-normality and heterogeneity in the non-orthogonal two-way ANOVA design," *British Journal of Mathematical and Statistical Psychology*, vol. 54, no. 1, pp. 79–94, 2001.

[19] Y. Q. Li, J. Zhou, and X. Z. Zhou, "Stage-Discharge Curve Fitting Based on Box-Cox Transform," *Journal of Yangtze River Scientific Research Institute*, vol. 37, 2018.

[20] J. Y. Yoon, J. Kim, and S. Song, "Comparison of parameter estimation methods for normal inverse Gaussian distribution," *Communications for Statistical Applications and Methods*, vol. 27, no. 1, pp. 97–108, 2020.

[21] V. M. Guerrero and R. A. Johnson, "Use of the box-cox transformation with binary response models," *Biometrika*, vol. 69, no. 2, pp. 309–314, 1982.

[22] N. Taylor, "Realised variance forecasting under Box-Cox transformations," *International Journal of Forecasting*, vol. 33, no. 4, pp. 770–785, 2017.

[23] M. J. Gurka, L. J. Edwards, K. E. Muller, and L. L. Kupper, "Extending the Box-Cox transformation to the linear mixed model," *Journal of the Royal Statistical Society: Series A*, vol. 169, no. 2, pp. 273–288, 2006.

[24] T. Hamasaki and S. Y. Kim, "Box and Cox power-transformation to confined and censored nonnormal responses in regression," *Computational Statistics & Data Analysis*, vol. 51, no. 8, pp. 3788–3799, 2007.

[25] T. Proietti and H. Lütkepohl, "Does the Box-Cox transformation help in forecasting macroeconomic time series?" *International Journal of Forecasting*, vol. 29, no. 1, pp. 88–99, 2013.

[26] E J J, "MINITAB: HandStudent book," *Technometrics*, vol. 20, no. 2, pp. 211-212, 1978.

[27] Q. Y. Zhou, P. Tian, and Z. Y. Tian, "Estimation of non-normal process capability index based on johnson system of transformations (in Chinese with English abstract)," *Systems Engineering*, vol. 22, pp. 98–102, 2004.

[28] J. F. Slifker and S. S. Shapiro, "The johnson system: selection and parameter estimation," *Technometrics*, vol. 22, no. 2, pp. 239–246, 1980.

[29] A. Mateus and M. Tome, "Estimating the parameters of the johnson's S-B distribution using an approach of method of moments," *Numerical Analysis and Applied Mathematics Icnaam 2011: International Conference on Numerical Analysis and Applied Mathematics*, vol. 1389, 2011.

[30] H. J. H. Tuenter, "An algorithm to determine the parameters ofS_U-curves in the johnson system of probabillity distributions by moment matching," *Journal of Statistical Computation and Simulation*, vol. 70, no. 4, pp. 325–347, 2001.

[31] I. D. Hill, R. Hill, and R. L. Holder, "Algorithm AS 99: fitting johnson curves by moments," *Applied Statistics*, vol. 25, no. 2, p. 180, 1976.

[32] Y. M. Chou, A. M. Polansky, and R. L. Mason, "Transforming non-normal data to normality in statistical process control," *Journal of Quality Technology*, vol. 30, no. 2, pp. 133–141, 1998.

[33] S. T. Mandraccia, G. D. Halverson, and Y. M. Chou, "Control chart design strategies for skewed data," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 2876, pp. 196–205, 1996.

[34] A. M. Polansky, Y. M. Chou, and R. L. Mason, "An algorithm for fitting Johnson transformations to non-normal data," *Journal of Quality Technology*, vol. 31, no. 3, pp. 345–350, 1999.

[35] P. F. Li, G. Yang, X. L. He, F. D. Li, K. Yan, and Z. L. Wang, "Effects of drip irrigation on components of water cycle in arid inland areas: a case study of Manas river basin in northwestern China," *International Journal of Agricultural and Biological Engineering*, vol. 12, no. 1, pp. 132–138, 2019.

[36] H. B. Ling, H. L. Xu, W. Shi, and Q. Q. Zhang, "Regional climate change and its effects on the runoff of Manas River, Xinjiang, China," *Environmental Earth Sciences*, vol. 64, no. 8, pp. 2203–2213, 2011.

[37] Z. M. Liang, Y. M. Hu, and J. Wang, "Advances in hydrological frequency analysis of non-stationary time series (in Chinese with English abstract)," *Advances in Water Science*, vol. 22, pp. 864–871, 2011.

[38] M. N. Khaliq, T. Ouarda, J. C. Ondo, P. Gachon, and B. Bobee, "Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review," *Journal of Hydrology*, vol. 329, no. 3-4, pp. 534–552, 2006.

[39] P. Xie, G. C. Chen, and J. Xia, "Hydrological frequency calculation principle of inconsistent annual runoff series under Changing environments (in Chinese with English abstract)," *Engineering Journal Of Wuhan University*, vol. 38, pp. 8–11+17, 2005.

[40] G. Q. Wu, "Application of social spider optimization algorithm in parameter optimization of hydrological frequency curve (in Chinese wih English abstract)," *Journal of Water Resources and Water Engineering*, vol. 26, pp. 123–126+131, 2015.

[41] J. M. Zhang, X. H. Chen, and C. Q. Ye, "Calculation of negative-skewness hydrological series with Pearson-III frequency curve (in Chinese with English abstract)," *Journal of Hydraulic Engineering*, vol. 43, pp. 1296–1301, 2012.

[42] M. I. Brunner, R. Furrer, and A. C. Favre, "Modeling the spatial dependence of floods using the Fisher copula," *Hydrology and Earth System Sciences*, vol. 23, no. 1, pp. 107–124, 2019.

[43] R. C. Phillips, S. Z. Samadi, and M. E. Meadows, "How extreme was the October 2015 flood in the Carolinas? An assessment of flood frequency analysis and distribution tails," *Journal of Hydrology*, vol. 562, pp. 648–663, 2018.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 7929449, 9 pages
https://doi.org/10.1155/2022/7929449

Research Article

Chinese Contemporary Music Diffusion Strategy Based on Public Opinion Maximization

Yuanyuan Yang

Xinxiang University, Xinxiang 453000, China

Correspondence should be addressed to Yuanyuan Yang; yangyuanyuan0527@xxu.edu.cn

Received 17 May 2022; Revised 10 June 2022; Accepted 14 June 2022; Published 28 June 2022

Academic Editor: Xingsi Xue

Chinese contemporary music has made tremendous progress in the twentieth century. After entering the twentieth century, with the rise of pop music, Chinese contemporary music has gradually been forgotten by people. Apparently, Chinese contemporary music has a long history and covers huge varieties. Nevertheless, the diffusion of Chinese contemporary music is restricted by many factors. The fundamental reason is that people's cognition of Chinese contemporary music has changed. However, the rise of social networks provides a new opportunity for the diffusion of Chinese contemporary music. By searching nodes with high influence in social networks, where the node is enabled by the sensor, we transform the problem for maximizing public opinion into a process of searching nodes set of maximum influence, thus providing power for the diffusion of Chinese contemporary music. In this paper, from the perspective of percolation, the network connectivity changes caused by node failure are considered, and the set selection problems of high-influence individuals are mapped to the optimal percolation problem when the network has a percolation transition. Then an influence maximization algorithm is proposed based on the network percolation process, and the efficiency of the proposed algorithm is optimized. Through experiments, we can infer the reasonable range of maximum searching times for the influence maximization algorithm. Meanwhile, compared with other baselines, the searching time of the algorithm proposed in this paper is also the least. The experimental results reveal that the proposed algorithm is feasible to diffuse Chinese contemporary music and maximize its public opinion and has a strong impetus for the revival of Chinese contemporary music.

1. Introduction

Any kind of music is inevitable in history. The creation of Chinese contemporary music reflects unique creative elements and concepts in the fast-changing development trend of today's society. With its unique national style and unique charm, it advances continuously, presenting the development trend of modernization and diversification [1–3]. The twentieth century is a multicultural century; the emergence of music is the inevitable result of historical development, so the twentieth-century music creation has also entered a complex, modernized, and diversified new era. In this century, expressionist music, sequential music, collage music, accidental music, electronic music, and many other schools appeared in the West [4, 5]. In the creation, the composer integrates the new music language and music creation techniques, and the new audio-visual music works emerge at the historic moment. Although China has embarked on a completely different combination of Chinese and Western traditional folk music since the beginning of the twentieth century, it was not until the 1970s and 1980s after the reform and opening-up that mainland China really caught up with the development of modern music. Previously, the Western theory of musical form structure has deeply influenced Chinese music theory. During the period of reform and opening-up, a number of musicians, music educators, and theorists in China fully absorbed Western composition technology theories, concentrated on studying Western music works, and widely learned and accepted Western instruments, bands, and music styles.

Since the 1980s, some composers have made breakthrough explorations of the direction, ideas, methods, and aesthetic concepts of Chinese contemporary music, which have played a positive role in promoting the development of

Chinese contemporary music [6–8]. Chinese contemporary music works seem to have to fill the gaps in many ways and levels. Composers have gradually extended to other artistic directions, such as applying folk music, jazz, and other music from different regions or cultures to their works, using various modern harmony, tone scale, rhythm, and tone color to varying degrees and combining these new materials and techniques with their own creative ideas to create refreshing and diverse styles of music works. To express the content, style, and needs of personalized language, composers freely choose a variety of musical genres.

To some extent, it can be said that the creation techniques of modern music originate from the West, and the musical language is different from the traditional music and aesthetic concepts in China [9]. However, the success of a work must be a perfect combination of various factors, such as times and distinctive national characteristics as well as personal style. In modern music creation, some composers have been exploring and searching for the means and methods to embody Chinese factors and strive to create modern music works with Chinese style, whose fundamental purpose is to create the Chinese atmosphere of the works and provide a "Chinese" background for the works [10–13]. And this Chinese factor is directly reflected in the choice of music creation themes. To pursue the goal of modern music and embody the essence of Chinese culture, composers often directly point out some connection between their works and China in the way of titles and strive to express the content with deep Chinese charm.

As a ritual and music nation with a long culture, the diversity of Chinese music has laid a foundation for the development of Chinese music creation both in content and form. Only by judging and sorting all kinds of data from a multidimensional perspective can the argument be more academic [14]. Under the impact of the new era background, the study of Chinese contemporary music shows strong characteristics of the era. The marketization research of music reflects the productive property of music after the transformation of China's contemporary market economy, and it becomes a commodity that can be profitable in the market. The study of music globalization is urgent to consider the future of Chinese music under the pattern of globalization. Especially, the impact of pop music on modern music makes the diffusion of Chinese contemporary music more and more difficult. Different from tape and CD in the past, today's social networks sweep the Internet, and people are closely connected with others all the time; the diffusion of music seems to become easy, which also provides an opportunity for the diffusion of Chinese contemporary music.

Relying on the Internet, social networks show a rapid development trend, and the advent of email opened the prelude to the development of social networks. Social network platforms such as WeChat and TikTok have shown a rapid development trend relying on the Internet. In social networks, everyone can share their ideas and opinions with others immediately, which not only enriches people's spare time but also promotes information exchange between people [15–17]. The speed at which information disseminates on social networks is so dramatic that a single piece of information can reach tens of thousands in a short time. As one of the most important problems in social network analysis, influence maximization intends to find high-impact users in the social network [18]. Identifying these high-impact users can provide decision-making assistance for the diffusion of Chinese contemporary music and inject vitality into Chinese contemporary music. The existing methods to measure user influence generally rely on the network topology or define the information propagation model to count the information propagation scope of users under the model, which is mostly based on static networks. However, real social networks tend to change dynamically over time, making it difficult for existing measures to accurately describe users' real influence. In the propagation of Chinese contemporary music, we hope to identify high-influence users in the network and diffuse its propagation.

We describe our contributions as follows:

(1) We apply the influence maximization to the propagation strategy of Chinese contemporary music and find the nodes with the maximum influence, so as to maximize public opinion and diffuse the influence of Chinese contemporary music.

(2) From the perspective of percolation, we consider the network connectivity changes caused by node fault and map the set selection problem of high-influence nodes in influence maximization to the optimized percolation problem.

(3) An influence maximization algorithm is proposed based on the robustness of the network percolation process, and the efficiency of the algorithm is optimized according to the monotonicity of the maximum connected cluster size in the percolation process.

The rest of the paper is structured as follows. In Section 2, we review the related works. An influence maximization algorithm based on site percolation is proposed in Section 3. Experimental results are presented in Section 4, and finally, Section 5 gives the conclusion of this paper.

2. Related Works

2.1. Artificial Intelligence and Music. As technology for simulating, extending, and expanding human intelligence, artificial intelligence has been advanced from the primary form of computational intelligence to the advanced form of perceptual intelligence and cognitive intelligence [19–22]. Music is one of the reappearance forms of human emotion and cognition, and artificial intelligence technology has gradually penetrated into every link of music diffusion. In the narrow sense, music information refers to the inner artistic information that is composed of a group of interrelated and meaningful musical symbols and can express some complete meaning. Generalized music information also includes external data information such as music carrier, music copyright, and music user preference. The participation of artificial intelligence in the process of music

information collection can be divided into two parts: one is the collection of physical music information [23–25]. For example, music information retrieval (MIR) extracts audio features based on audio signal processing and extracts music information from the audio content. The other is the collection of music external carrier, copyright, and user preference information.

AI music generation platform uses genetic algorithm [26], neural network [27], Markov chain [28], and hybrid algorithm [29] technology to make rules for the computer, in the process of deep learning of large-scale music library analysis of composition rules and other music information for re-creation, so as to achieve independent composition, lyrics, and other music creation functions.

2.2. Influence Maximization. Music should be shared, as well as Chinese contemporary music. With the popularization of pop music, Chinese contemporary music is gradually forgotten. How to share Chinese contemporary music with more people and maximize the diffusion of influence is the main content of influence maximization. Information in social networks is generated by individuals and then disseminated along the edges between individuals. Influence refers to the phenomenon that individuals' opinions and information on the network influence others and eventually form a cascade of information transmission. In [30], the authors introduced the issue of maximizing multiple influences across multiple social networks, that is, influential users could accept multiple products for free, while influential users had sufficient purchasing power to use multiple promotions in their social interactions. The existing influence diffusion model does not consider overexposure; in [31], the authors proposed the influence diffusion model to capture overexposure and studied the problem of maximum influence under the independent cascade model of delay perception with overexposure. In [32], the authors presented an effective method to maximize the influence in signed networks, so as to maximize the positive influence diffusion in signed networks. In [33], the authors proposed a problem of maximizing efficiency considering the influence of diffusion delay. In [34], the authors proposed a practical algorithm for adaptive influence maximization and a general framework that can be instantiated by any existing nonadaptive influence maximization algorithm with expected approximation guarantees.

However, to the best of our knowledge, few scholars have studied the diffusion strategies of Chinese contemporary music. Therefore, we introduce the maximization of influence into the diffusion strategy to find the nodes set with the biggest influence, so as to maximize public opinion and diffuse the Chinese contemporary music.

3. Influence Maximization Algorithm Based on Site Percolation

The removal of some nodes and the edges connected to them from the network is called percolation, more accurately, site percolation. Percolation can be used as a model for many real-world phenomena, such as router failures on the Internet, traffic jams in urban road networks, and species extinction in the food web. One of the purposes of studying percolation is how node failure affects the whole network operation. It is very important to identify the nodes that have a great influence on the function of the network system. By diffusing the influence range of these nodes, Chinese contemporary music can be widely replayed.

Most nodes in the network are connected to form a large branch without removing or removing only a small number of nodes. As percolation proceeds, when the removal ratio of nodes reaches a certain value, the network will split into several small branches, which is called percolation transition [35]. The percolation threshold is the core content of percolation research and is of great significance for network control and network robustness research. Percolation provides a natural model for the study of network robustness, but the percolation threshold is not the only important parameter. Because most nodes of the network have failed when the percolation threshold is approached, another important aspect of studying percolation is to observe how the maximum connected subgraph size of the remaining network changes with the removal of nodes. The size of the maximum connected subgraph indicates that at least most of the networks are in normal operation. Therefore, this subsection selects the relative size of the maximum connected subgraph to study the robustness of the network.

3.1. Problem Definition. The problem of this paper can be specifically described as how to remove nodes in the network to minimize the robustness of the remaining network, that is, to maximize the impact on the structure and function of the network. The idea is that the robustness of the percolation algorithm is affected by the relative size of the maximum connected subgraph of the network. To minimize the robustness of the network, the nodes added each time should minimize the growth rate of the maximum connected cluster size of the network in the process of restoring the entire network from the empty network, while how to make the largest connected cluster in the network grow slowly becomes our concern. To merge clusters and update the maximum connected cluster size, two conditions need to be met: (i) the node to be added has a neighbor node and the neighbor node has been added to the network and (ii) the cluster number of the node to be added is different from that of the neighbor node.

If updating the maximum connected cluster size is to be avoided as much as possible, it can be obtained from condition (i) that the number of neighbor nodes of the added node should be as small as possible or the neighbor nodes have not been added to the network. According to condition (ii) and the cluster combination rule, the cluster number of all neighbor nodes of the added node should be the same as far as possible. To sum up, it can be inferred from the above that the node selection rule of recovering the whole network from an empty network is as follows: (a) the number of neighbor nodes should be as small as possible and (b) the neighbor node has not been added to the network or the

Input: $G(V, E)$, Q
Output: Q_{opt}
(1) initialize $R = R(V, E)$, $V(R) \leftarrow \varnothing$, $E(R) \leftarrow \varnothing$
(2) $Q_{\text{opt}} \longleftarrow \varnothing$
(3) **while** $Q \neq \varnothing$ **do**
(4) find $u \leftarrow \arg\min d(V(R) \cup \{v\})$
(5) add node u and its corresponding edges to network R
(6) Dequeue (Q, U)
(7) Enqueue (Q_{opt}, U)
(8) **end while**
(9) return Q_{opt}

Algorithm 1: Influence maximization algorithm based on site percolation.

Input: $G(V, E)$, Q
Output: u
(1) $k \longleftarrow 0$
(2) sort queue Q by $d(V \cup \{v\})$
(3) **for** $k \leq m$ **do**
(4) $v_k = k$th node in Q
(5) calculate $d(U \cup \{v_k\})$ and update $\mathrm{d}(U \cup \{v\})_{\min}$
(6) $k = k + 1$
(7) **if** $d(U \cup \{v\})_{\min} \leq d(V \cup \{v_{k+1}\})_{\min}$
(8) break
(9) **end if**
(10) **end for**
(11) return u in $d(U \cup \{v\})_{\min}$

Algorithm 2: Node optimal search algorithm.

cluster number of all neighbor nodes should be the same as possible.

Rule (a) indicates that the centrality measure most directly related to the growth rate of the maximum connected cluster size is the degree of nodes. The higher the degree of the node is, the faster the size of the maximum connected cluster grows after the node is added. Therefore, an initial queue Q can be constructed according to the degree of the node. Each time, a node with the lowest degree is selected from the network and added to the queue Q, and each node added must be removed from the network, and all remaining nodes are reordered according to the degree value until all nodes are added to the queue Q. Then, according to rule (b), a node that minimizes the change of the largest connected cluster size is selected from the sequence every time.

3.2. Influence Maximization Algorithm. Let $N(R)$ represent the node set in network R and $C(N)$ represent the size of the largest connected cluster in network R. Given social network $G(V, E)$, where V represents the set of nodes and E represents the set of edges. How to find K initial users so that information can be disseminated through these initial users and the number of affected users can reach the maximum, that is

Table 1: Experimental data sets.

	socfb-Caltech36	socfb-UC61	Rovira i Virgili	WordNet
N	796	13,746	1,133	146,005
M	16,656	442,174	5,451	656,999
K	43.32	64.33	9.62	9.001
C	0.408	0.261	0.165	0.097

$$D' = \arg\max \sigma(D), \tag{1}$$

where D represents the initial user set, $\sigma(D)$ represents the number of affected users, $D \subseteq V$, and $|D| = K$.

The influence maximization algorithm is shown in Algorithm 1.

Assuming that the time complexity for adding nodes is $O(N)$ and the time complexity for adding edges is $O(M)$. The most time-consuming part of Algorithm 1 is the process of finding the maximum connected cluster size with the minimum change. Every time an optimal node u is found in the whole algorithm process, its $d(V \cup \{v\})$ is no longer calculated in the subsequent search process. The whole process is equivalent to make $N/2$ percolation, so the time complexity of the whole algorithm is $O(N^2 \log N + \mathrm{NM})$. The whole process calculates $N(N + 1)/2$ $\mathrm{d}(V \cup \{v\})$ twice in

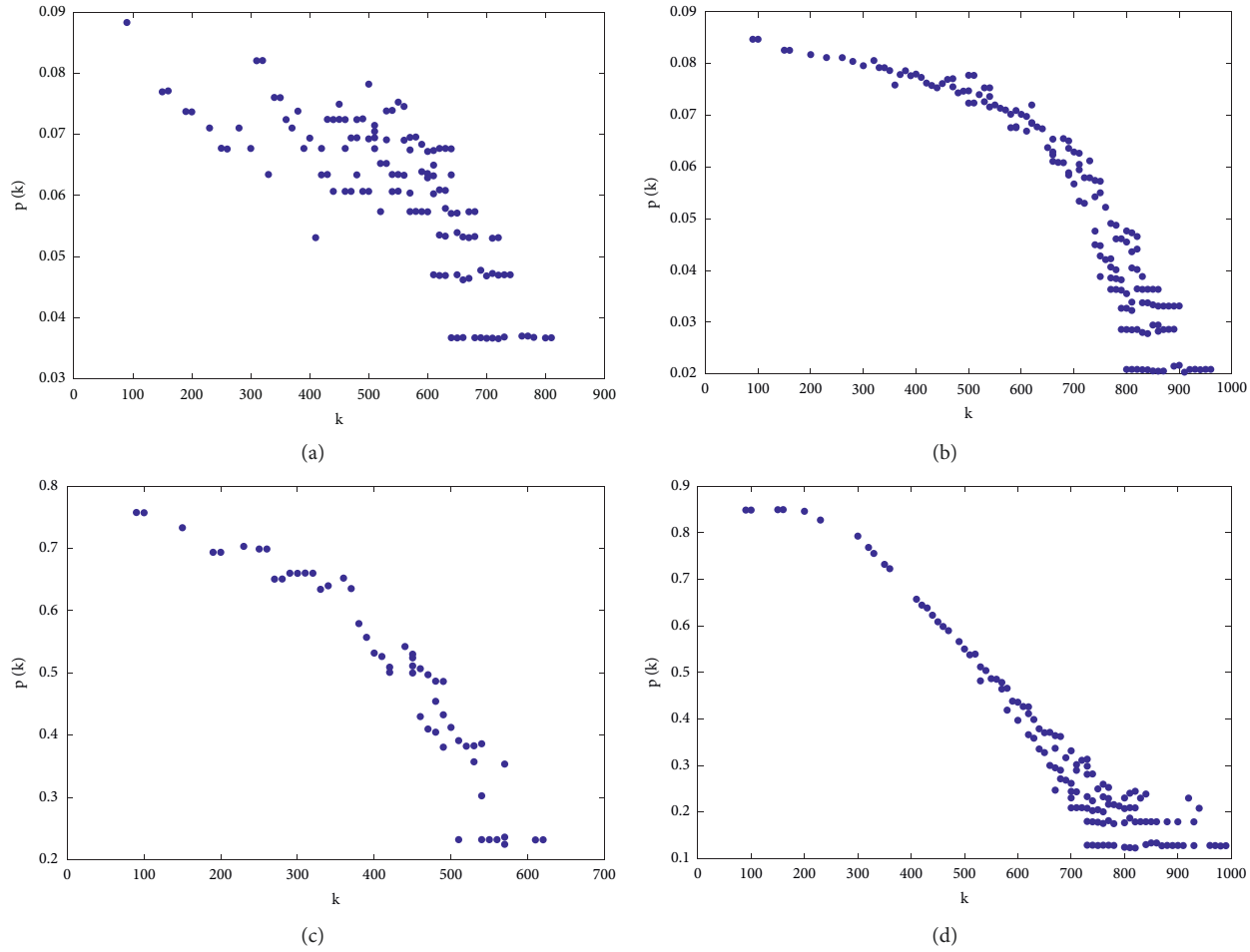


Figure 1: Node degree distribution of four data sets: (a) socfb-caltech36 data set, (b) socfB-UC61 data set, (c) Rovira i Virgili data set, and (d) WordNet data set.

total, and the average time complexity of calculating $d(V \cup \{v\})$ is $O(\log N + M/N)$.

3.3. Optimization of Algorithm Efficiency. When nodes are added to network R, the size of the maximum connected cluster either increases or is unchanged. Apparently, $d(V)$ is monotonically increasing and non-negative, that is

$$d(V \cup \{v\}) \geq d(V). \tag{2}$$

Then the following inequality is easily proved to be true:

$$\forall V \subset U \subset N, \forall v \in N, v \notin U, d(V \cup \{v\}) \leq d(U \cup \{v\}). \tag{3}$$

Equation (3) can be used to draw the following conclusions.

The size of the largest connected cluster in the network after adding node v can only increase or remain the same with each iterative calculation for any node v, which means that the size of the largest connected cluster in the network after adding node v in this iterative calculation can be used as the lower limit of the next iterative calculation.

To find the node in the queue that minimizes the change of the largest connected cluster size, the algorithm can be reduced, and the search efficiency can be improved by applying the above conclusions and setting the maximum search times. The specific implementation method is as follows.

Before each iterative calculation, the calculation results of the previous iteration shall be sorted from small to large as the lower limit of the iterative calculation, and the nodes shall be evaluated in this order. For each node evaluated, the minimum value of the maximum connected scale of the network must be updated. If the minimum of the maximum network connected scale after a node is added is less than or equal to the lower value of the next node to be evaluated or reaches the maximum search times, the search can be stopped. Since under the premise of not exceeding the maximum search times, the node that causes the minimum change of the maximum connected scale of the network has been found.

The node optimal search algorithm is shown in Algorithm 2. The most time-consuming part of the algorithm is to calculate the maximum network connected size $d(U \cup \{v_k\})$ after node v_k is added. According to Algorithm 1, the average time complexity of $d(U \cup \{v_k\})$ is $O(\log N + M/N)$. Since the number of searching optimal nodes per time does not exceed m, the average time complexity of the optimal search algorithm does not exceed

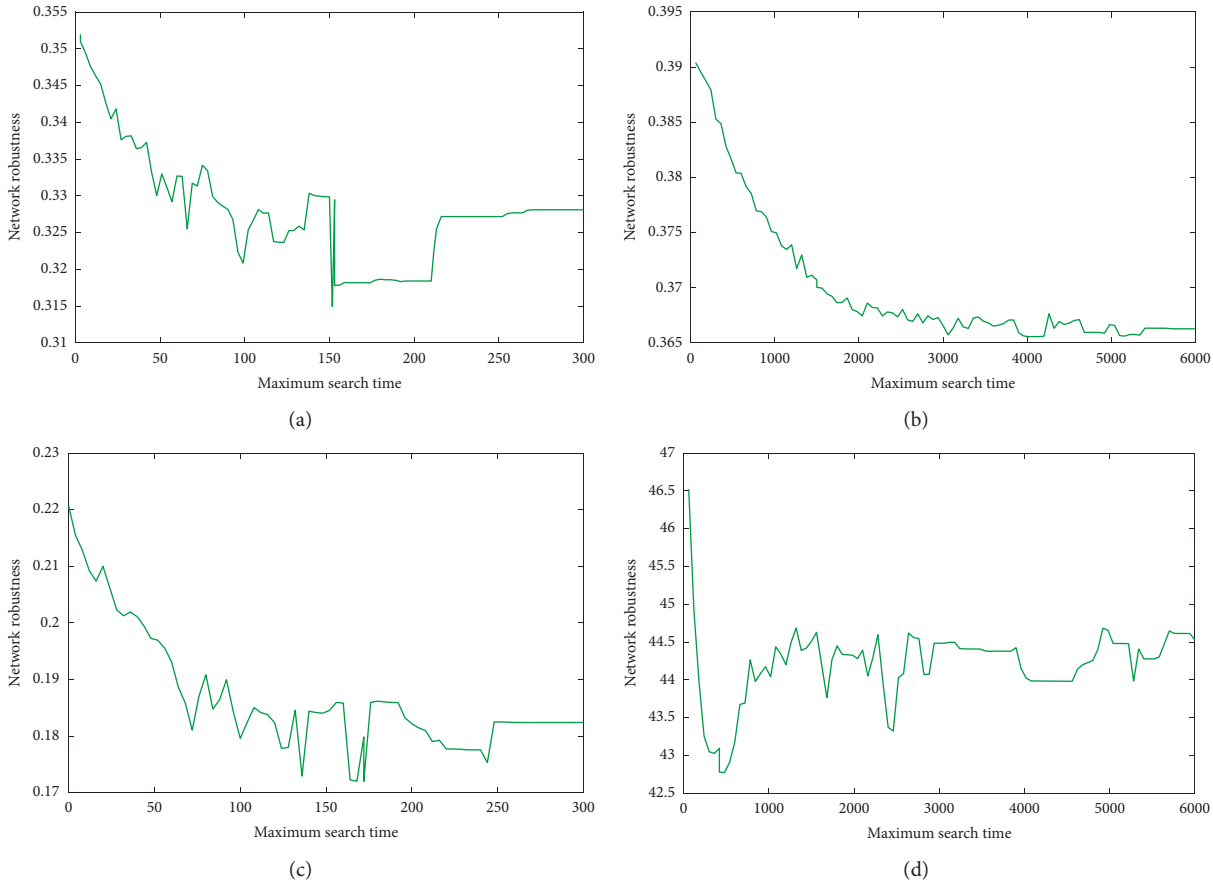


Figure 2: Change of network robustness with a maximum number of search times: (a) socfb-caltech36 data set, (b) socfB-UC61 data set, (c) Rovira i Virgili data set, and (d) WordNet data set.

$O(m \log N + \mathrm{mM}/N)$, and the average time complexity of the influence maximization algorithm does not exceed $O(m \log N + \mathrm{mM}/N)$.

4. Experiments

4.1. Setup. The experimental data set and its basic attributes are shown in Table 1, where N is the total number of nodes in the network, E is the total number of edges in the network, K is the average degree of the network, and C is the average clustering coefficient. Socfb-caltech36 and SocfB-UC61 data sets in Table 1 are social networks extracted from Facebook, where nodes in the network represent users, and edges represent friend relationships between users. Rovira i Virgili is the e-mail communication network of the University of Tarragona in Spain, where nodes in the network represent users and edges indicate that at least one e-mail has been sent between users. WordNet is the vocabulary network in the WordNet data set, where nodes in the network represent English words and edges represent the relationships between them, such as synonyms and antonyms [36]. In real life, there are many large-scale complex networks, such as Internet, social networks, transportation network, biological network, and so on. In the beginning, the degree distribution of these complex networks is Poisson distribution. For social networks, the degree of each node obeys the power law distribution. Most common nodes have few connections, and a few popular nodes have many connections. Such networks are called scale-free networks. We plotted the node degree distributions of the above four data sets to observe their scale-free characteristics. The node degree distribution is shown in Figure 1, where $p(k)$ is the proportion of nodes with degree k in the entire network. The data of socfb-Caltech36 and Rovira i Virgili data sets are relatively sparse, resulting in insignificant scale-free characteristics. The tail of socfB-UC61 and WordNet data sets generally obey the power law distribution, and their power law index is between 2 and 3, showing obvious scale-free characteristics. Therefore, we selected two obvious scale-free data sets and two non-obvious scale-free data sets as experimental data sets.

4.2. Algorithms in Comparison. Since the effect of the influence maximization algorithm is affected by the maximum number of searches m, different m is selected for further test on the four data sets to observe the influence of the maximum number of searches m on the network robustness. Experimental results of the network robustness of the four data sets varying with the maximum search times m are shown in Figures 2(a)–2(d). The curve in the figure represents the variation curve of network robustness with the maximum search times m. It can be seen from Figures 2(a)–

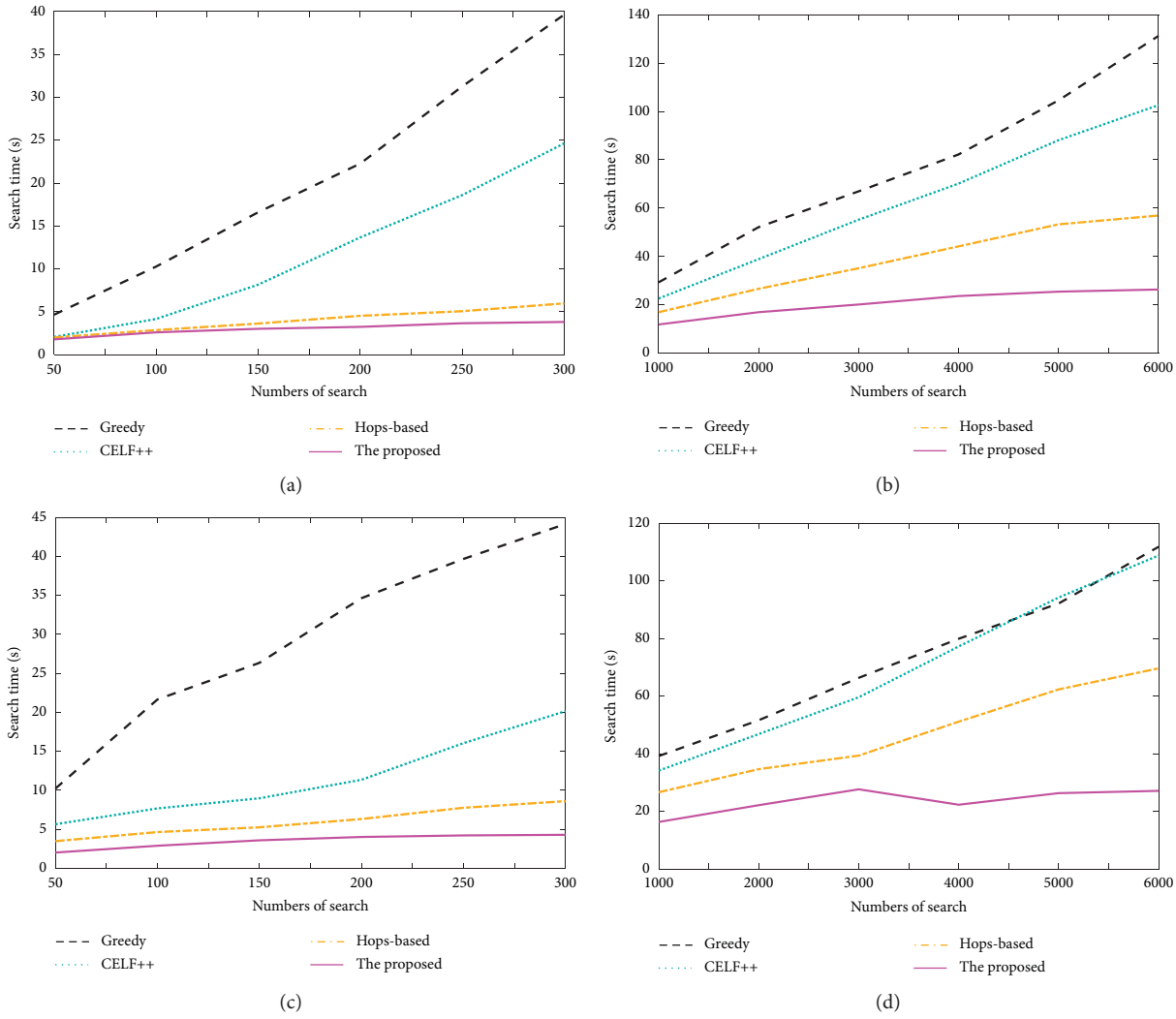


Figure 3: Search time: (a) socfb-caltech36 data set, (b) socfB-UC61 data set, (c) Rovira i Virgili data set, and (d) WordNet data set.

2(d) that with the increase of the maximum search times m, the network robustness tends to decline and becomes stable after a certain degree of decline. Therefore, we can infer that the reasonable value range of parameter maximum search times m of the influence maximization algorithm is between $[0.1B, 0.2B]$, where B is the number of nodes in the network.

Search time is another important metric to verify the effectiveness of the algorithm and plays an important role in the diffusion of Chinese contemporary music. We compare search time with three classical influence maximization algorithms, namely greedy algorithm, CELF++ algorithm, and Hop-based influence estimation algorithm. As can be seen from Figures 3(a)–3(d), the comparison of search time on the four data sets shows that with the increase in search times, the influence maximization algorithm based on site percolation proposed in this paper has the slowest increase in search time and gradually converge after a certain number of times, and the search time is always the least. This is because the introduction of percolation reduces the search time by about half. The greedy algorithm can give a better solution set, but its time complexity is too high. CELF++ algorithm rearranges the nodes according to their size before each reassessment of the node's marginal gain. Each node is evaluated, and the maximum marginal gain is updated. However, Monte Carlo simulation is still used when evaluating the node's marginal gain. A large amount of time consumption makes it difficult to apply to large-scale social networks, so the search time is also longer.

5. Conclusion

Although Chinese contemporary music has been gradually forgotten after a hundred years of development, the rise of social networks provides a new opportunity for its diffusion. We design an influence maximization algorithm based on percolation mode and optimize the efficiency of the algorithm. Since the scale-free network has robustness under random failure, the problem of set selection of high-influence individuals in influence maximization can be transformed into a problem of network percolation, and the

advantages and disadvantages of the selected nodes set can be evaluated with network robustness, thus pointing out the direction for the diffusion of Chinese contemporary music. To minimize the robustness of the network, the node that increases the scale of the current maximum connected subgraph slowest can be selected each time in the percolation process. Therefore, the more backward the node is, the higher its influence will be. In the process of nodes selection, the searching efficiency of the proposed algorithm can be improved by increasing the size of the maximum connected subgraph. The experimental results reveal that the proposed influence maximization algorithm based on site percolation can search the nodes set of influence maximization with very little search time, which has strong theoretical support for the diffusion of Chinese contemporary music.

Although this paper has made some progress in the study of the maximization of influence of Chinese contemporary music diffusion, it is undeniable that there are still some deficiencies in the work. Several directions in the future are still worthy of further exploration and research:

(1) As the scale of data in different time periods is generally different and it is difficult to ensure the comprehensiveness of data during data collection, the length and quantity of network snapshots will have a certain influence on experimental results. To avoid or reduce the interference of these non-experimental factors, more analysis on data set processing is needed in future work.

(2) The main reason why the calculation results of the existing information diffusion model differ greatly from the real social network is that the probability of information diffusion among users depends on many uncertain factors such as people's interests and intimacy between people, and the influence of users in different fields and events is also different. Therefore, in order to truly reveal the influence of users, it is necessary to further study the probability of information diffusion among users.

(3) Chinese contemporary music has a long history and is an important part of Chinese culture. With the rapid development of the economy and science and technology, all walks of life have encountered unprecedented opportunities and challenges. In the era of new media, the diffusion and development of Chinese contemporary music have also faced new changes. Under such circumstances, Chinese contemporary music has encountered development bottlenecks both in content and form but has new opportunities in diffusion. Excellent Chinese contemporary music can make use of the characteristics of the times and new media to promote its vigorous development. The diffusion path of Chinese contemporary music in the new media era is characterized by diversification. In view of the current situation of the development of Chinese contemporary music, we should take advantage of the opportunities brought by new media and adopt various ways to diffuse Chinese contemporary music around the world.

Data Availability

All data used to support the findings of the study are included within this article.

Conflicts of Interest

The author declares that there are no conflicts of interest in this article.

Acknowledgments

This work was supported by Education Department Project of Henan Province: "A Study on the Communication of Chinese Contemporary Musical Culture along the Maritime Silk Road" (Grant no. 2020GGJS252).

References

[1] C. Wei, "The inheritance of Chinese traditional music in college music education,"vol. 53, pp. 167–171, in *Proceedings of the 2017 7TH INTERNATIONAL CONFERENCE ON EDUCATION AND MANAGEMENT (ICEM 2017)*, vol. 53, pp. 167–171, Atlantis Press, Netherlands, December 2017.

[2] C. Chen, "The Comparison between Chinese and western music education,"vol. 54, pp. 977–980, in *Proceedings of the 2016 INTERNATIONAL CONFERENCE ON EDUCATION, SPORTS, ARTS AND MANAGEMENT ENGINEERING*, vol. 54, Atlantis Press, Netherlands, March 2016.

[3] N. Song, "The Development Situation of Chinese National Music Teaching," in *Proceedings of the 2017 INTERNATIONAL CONFERENCE on ADVANCED EDUCATION, PSYCHOLOGY AND SPORTS SCIENCE (AEPSS)*, pp. 86–90, Chinese national music, music teaching, development, China, 2017.

[4] J.-P. Després, P. Burnard, F. Dubé, and S. Stévance, "Expert western classical music improvisers' strategies," *Journal of Research in Music Education*, vol. 65, no. 2, pp. 139–162, 2017.

[5] W. Hao, "Research on the Status Quo of the Study of Western Music History in China," in *Proceedings of the 2018 7TH INTERNATIONAL CONFERENCE on SOCIAL SCIENCE, EDUCATION AND HUMANITIES RESEARCH (SSEHR 2018)*, pp. 161–164, Music College, Baotou Teachers' College, Baotou China, December 2018.

[6] H. Wang and J. Wang, "The deficiency and embarrassment of Chinese music history, 2014 2ND international conference on social science and health,"vol. 58, pp. 105–108, in *Proceedings of the 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018)*, vol. 58, pp. 105–108, Atlantis Press, Netherlands, November 2018.

[7] X. Zheng, D. Li, L. Wang, L. Shen, Y. Gao, and Y. Zhu, "An automatic composition model of Chinese folk music," *ADVANCES IN MATERIALS, MACHINERY, ELECTRONICS I*, vol. 1820, 2017.

[8] X. Hu and J. H. Lee, "Towards global music digital libraries," *Journal of Documentation*, vol. 72, no. 5, pp. 858–877, 2016.

[9] L. Tan, "Towards an ancient Chinese-inspired theory of music education," *Music Education Research*, vol. 18, no. 4, pp. 399–410, 2016.

[10] S. Petersen, "Camp, the musical self-concept of Chinese music students," *Frontiers in Psychology*, vol. 7, 2016.

[11] L. L. Lyu, "The creative application of computer technology in Chinese traditional music culture,"vol. 116, pp. 268–274, in *Proceedings of the INTERNATIONAL CONFERENCE ON COMMUNICATION AND ELECTRONIC INFORMATION ENGINEERING (CEIE 2016)*, vol. 116, Atlantis Press, Netherlands, October 2017.

[12] R. Li, "Chinese Folk Music: Study and Dissemination through Online Learning Courses," *EDUCATION AND INFORMATION TECHNOLOGIES*, vol. 27, 2022.

[13] Y. Wang, Y. Li, and T. Chen, "The interactive design about Chinese classical music based on E-learning," in *Proceedings of the 2017 8TH INTERNATIONAL CONFERENCE ON E-EDUCATION, E-BUSINESS, E-MANAGEMENT AND E-LEARNING (IC4E 2017)*, pp. 46–50, Association for Computing Machinery, New York, NY, USA, January 2015.

[14] J. Zou, "Humanistic cultural tradition background and research of similarities and differences between Chinese and western music,"vol. 28, pp. 413–417, in *Proceedings of the 2014 2ND INTERNATIONAL CONFERENCE IN HUMANITIES, SOCIAL SCIENCES AND GLOBAL BUSINESS MANAGEMENT (ISSGBM 2014)*, vol. 28, pp. 413–417, School of Music Xianyang Normal University, Xianyang, China, December 2014.

[15] J. Li and G. Yang, "Network embedding enhanced intelligent recommendation for online social networks," *Future Generation Computer Systems*, vol. 119, pp. 68–76, 2021.

[16] E. B. Smith, R. A. Brands, M. E. Brashears, and A. M. Kleinbaum, "Social networks and cognition," *Annual Review of Sociology*, vol. 46, no. 1, pp. 159–174, 2020.

[17] C. Grange, I. Benbasat, and A. Burton-Jones, "A network-based conceptualization of social commerce and social commerce value," *Computers in Human Behavior*, vol. 108, 2020.

[18] H. Huang, H. Shen, and Z. Meng, "Community-based influence maximization in attributed networks," *Applied Intelligence*, vol. 50, no. 2, pp. 354–364, 2020.

[19] X. Du, "Application of deep learning and artificial intelligence algorithm in multimedia music teaching," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 6, pp. 7241–7251, 2020.

[20] K. Balasubramanian, "Symmetry, combinatorics, artificial intelligence, music and spectroscopy," *SYMMETRY-BASEL*, vol. 13, no. 10, 2021.

[21] B. Zhang and J. Lin, "Lecture Notes in Real-Time Intelligent Systems,"vol. 613, pp. 137–144, in *Proceedings of the Elements of Music Based on Artificial Intelligence, LECTURE NOTES IN REAL-TIME INTELLIGENT SYSTEMS (RTIS 2016)*, vol. 613, pp. 137–144, springer, Salmon, September 2016.

[22] S. Wu, "Research on the automatic music creation system based on the human-computer interaction and artificial intelligence,"vol. 55, pp. 640–645, in *Proceedings of the 2016 2ND INTERNATIONAL CONFERENCE ON SOCIAL SCIENCE AND TECHNOLOGY EDUCATION (ICSSTE 2016)*, vol. 55, Atlantis Press, Netherlands, May 2016.

[23] J. S. Seo, "A local feature aggregation method for music retrieval," *IEICE - Transactions on Info and Systems*, vol. E101.D, no. 1, pp. 64–67, 2018.

[24] A. Xambó, A. Lerch, and J. Freeman, "Music information retrieval in live coding: a theoretical framework," *Computer Music Journal*, vol. 42, no. 4, pp. 9–25, 2018.

[25] M. Weinberger and D. Bouhnik, "Various information aspects following the emergence of music streaming applications," *Online Information Review*, vol. 45, no. 1, pp. 118–137, 2021.

[26] B. Stoltz and A. Aravind, "MU_PSYC: Music Psychology Enriched Genetic Algorithm," in *Proceedings of the 2019 IEEE CONGRESS ON EVOLUTIONARY COMPUTATION (CEC)*, pp. 2121–2128, IEEE, Wellington, New Zealand, July 2019.

[27] J. A. Franklin, "Recurrent neural networks for music computation," *INFORMS Journal on Computing*, vol. 18, no. 3, pp. 321–338, 2006.

[28] Z. Hassani and A. I. Wuryandari, "Music generator with Markov chain: a case study with beatme touchdown," in *Proceedings of the 2016 6TH INTERNATIONAL CONFERENCE ON SYSTEM ENGINEERING AND TECHNOLOGY (ICSET)*, pp. 179–183, IEEE, Bandung, Indonesia, October 2016.

[29] R. Fox and R. Crawford, "A hybrid approach to automated music composition," *Advances in Intelligent Systems and Computing IN INTELLIGENT SYSTEMS*, vol. 464, pp. 213–223, 2016.

[30] S. S. Singh, K. Singh, A. Kumar, and B. Biswas, "MIM2: multiple influence maximization across multiple social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 526, 2019.

[31] G. Loukides, R. Gwadera, and S. W. Chang, "Overexposure-aware influence maximization," *ACM Transactions on Internet Technology*, vol. 20, no. 4, 2020.

[32] W. Liu, X. Chen, B. Jeon, L. Chen, and B. Chen, "Influence maximization on signed networks under independent cascade model," *Applied Intelligence*, vol. 49, no. 3, pp. 912–928, 2019.

[33] X. Zhu, Z. Wang, Y. Yang, B. Zhou, and Y. Jia, "Influence efficiency maximization: how can we spread information efficiently?" *JOURNAL OF COMPUTATIONAL SCIENCE*, vol. 28, pp. 245–256, 2018.

[34] K. Huang, J. Tang, K. Han et al., "Efficient approximation algorithms for adaptive influence maximization," *The VLDB Journal*, vol. 29, no. 6, pp. 1385–1406, 2020.

[35] M. Siomau, "Percolation transition control in quantum networks," *Quantum Information Processing*, vol. 19, no. 5, 2020.

[36] P. Kumaran and S. Chitrakala, "Topic Adaptive Sentiment Classification Based Community Detection for Social Influential Gauging in Online Social Networks," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 81, 2022.