Complexity

Deep Structure Representation and Learning for Complex Information Networks

Lead Guest Editor: Jia Wu
Guest Editors: Chuan Zhou, Shenghua Liu, and Haishuai Wang

Deep Structure Representation and Learning for Complex Information Networks

Complexity

Deep Structure Representation and Learning for Complex Information Networks

Lead Guest Editor: Jia Wu
Guest Editors: Chuan Zhou, Shenghua Liu, and Haishuai Wang

Contents

Prediction of circRNA-miRNA Associations Based on Network Embedding
Wei Lan, Mingrui Zhu, Qingfeng Chen, Jianwei Chen, Jin Ye, Jin Liu, Wei Peng, and Shirui Pan
Research Article (10 pages), Article ID 6659695, Volume 2021 (2021)

A Novel Emerging Topic Identification and Evolution Discovery Method on Time-Evolving and Heterogeneous Online Social Networks
Xiaoyan Xu, Wei Lv, Beibei Zhang, Shuaipeng Zhou, Wei Wei, and Yusen Li
Research Article (14 pages), Article ID 8859225, Volume 2021 (2021)

Anonymous Authentication and Key Agreement Scheme Combining the Group Key for Vehicular Ad Hoc Networks
Mei Sun, Yuyan Guo, Dongbing Zhang, and MingMing Jiang
Research Article (13 pages), Article ID 5526412, Volume 2021 (2021)

Multi-channel Convolutional Neural Network Feature Extraction for Session Based Recommendation
Zhenyan Ji, Mengdan Wu, Yumin Feng, and José Enrique Armendáriz Íñigo
Research Article (10 pages), Article ID 6661901, Volume 2021 (2021)

Graph-Based Analysis of RNA Secondary Structure Similarity Comparison
Lina Yang, Yang Liu, Xiaochun Hu, Patrick Wang, Xichun Li, and Jun Wu
Research Article (15 pages), Article ID 8841822, Volume 2021 (2021)

Multiplex Network Embedding Model with High-Order Node Dependence
Nianwen Ning, Qiuyue Li, Kai Zhao, and Bin Wu
Research Article (18 pages), Article ID 6644111, Volume 2021 (2021)

No-Reference Stereoscopic Image Quality Assessment Based on Binocular Statistical Features and Machine Learning
Peng Xu, Man Guo, Lei Chen, Weifeng Hu, Qingshan Chen, and Yujun Li
Research Article (14 pages), Article ID 8834652, Volume 2021 (2021)

A Novel Chinese Entity Relationship Extraction Method Based on the Bidirectional Maximum Entropy Markov Model
Chengyao Lv, Deng Pan, Yaxiong Li, Jianxin Li, and Zong Wang
Research Article (8 pages), Article ID 6610965, Volume 2021 (2021)

Vehicle Type Recognition Algorithm Based on Improved Network in Network
Erxi Zhu, Min Xu, and De Chang Pi
Research Article (10 pages), Article ID 6061939, Volume 2021 (2021)

Novel Node Centrality-Based Efficient Empirical Robustness Assessment for Directed Network
Xiaolong Deng, Hao Ding, Yong Chen, Cai Chen, and Tiejun Lv
Research Article (14 pages), Article ID 8715619, Volume 2020 (2020)

Joint Nonnegative Matrix Factorization Based on Sparse and Graph Laplacian Regularization for Clustering and Co-Differential Expression Genes Analysis
Ling-Yun Dai, Rong Zhu, and Juan Wang
Research Article (10 pages), Article ID 3917812, Volume 2020 (2020)

Deep Interest-Shifting Network with Meta-Embeddings for Fresh Item Recommendation
Zhao Li, Haobo Wang, Donghui Ding, Shichang Hu, Zhen Zhang, Weiwei Liu, Jianliang Gao, Zhiqiang Zhang, and Ji Zhang
Research Article (13 pages), Article ID 8828087, Volume 2020 (2020)

Maintenance and Operation Optimization Algorithm of PV Plants under Multiconstraint Conditions
Chi Hua, Liang Kuang, and Dechang Pi
Research Article (8 pages), Article ID 7975952, Volume 2020 (2020)

A Novel Expert Finding System for Community Question Answering
Nan Zhao, Jia Cheng, Nan Chen, Fei Xiong, and Peng Cheng
Research Article (8 pages), Article ID 5346085, Volume 2020 (2020)

Hindawi
Complexity
Volume 2021, Article ID 6659695, 10 pages
https://doi.org/10.1155/2021/6659695

Research Article

Prediction of circRNA-miRNA Associations Based on Network Embedding

Wei Lan,[1,2] **Mingrui Zhu**,[1] **Qingfeng Chen**,[1] **Jianwei Chen**,[1] **Jin Ye**,[1] **Jin Liu**,[3] **Wei Peng**,[4] **and Shirui Pan**[5]

[1]*School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China*
[2]*Guangxi Key Laboratory of Multimedia Communications and Network Technology, Guangxi University, Nanning 530004, China*
[3]*Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, China*
[4]*Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650050, China*
[5]*Faculty of Information Technology, Monash University, Melbourne 3800, Australia*

Correspondence should be addressed to Qingfeng Chen; qingfeng@gxu.edu.cn

Received 24 October 2020; Revised 7 January 2021; Accepted 20 August 2021; Published 15 September 2021

Academic Editor: Hassan Zargarzadeh

circRNA is a novel class of noncoding RNA with closed-loop structure. Increasing biological experiments have shown that circRNAs play an important role in many diseases by acting as a miRNA sponge to indirectly regulate the expression of miRNA target genes. Therefore, predicting associations between circRNAs and miRNAs can promote the understanding of pathogenesis of disease. In this paper, we propose a new computational method, NECMA, based on network embedding to predict potential associations between circRNAs and miRNAs. In our method, the Gaussian interaction profile (GIP) kernel similarities of circRNA and miRNA are calculated based on the known circRNA-miRNA associations, respectively. Then, the circRNA-miRNA association network, circRNA GIP kernel similarity network, and miRNA GIP kernel similarity network are utilized to construct the heterogeneous network. Furthermore, the network embedding algorithm is used to extract potential features of circRNA and miRNA from the heterogeneous network, respectively. Finally, the associations between circRNAs and miRNAs are predicted by using neighborhood regularization logic matrix decomposition and inner product. The performance of NECMA is evaluated by using ten-fold cross-validation. The results show that this method has better prediction accuracy than other state-of-the-art methods.

1. Introduction

circRNA is a new group of endogenous noncoding RNA that is highly represented in the mammalian transcriptome [1]. Compared with other noncoding RNAs (such as lncRNAs and miRNAs), circRNAs did not receive extensive attention in the early stage. With the development of high-throughput biological sequencing technology, more and more circRNA molecules have been discovered. Increasing studies have shown that circRNA does not have 5′-terminal cap and 3′-terminal poly (*A*) tail and can form a closed ring structure with covalent bonds [2]. Like other noncoding RNAs, circRNAs are also widely found in eukaryotes' brains [3], stomachs [4], and mammary glands [5]. Meanwhile, circRNAs are more stable than other linear noncoding RNAs due to their unique circular structure [6]. In addition, the unique structure of circRNA enables it to regulate gene transcription and expression [7]. For example, ciRs7 can bind related miRNAs and act as a miR-7 sponge to affect miR-7 binding to the target gene [8]. In addition, it has been found that circHIPK3 can sponge miR-124 and inhibit the activity of miR-124 in malignant tumors to achieve the purpose of regulating cell growth [9]. Numerous evidences have shown that miRNAs are closely related to a variety of

diseases [10]. For example, miR-145 inhibits colon cancer cell growth by targeting the insulin receptor substrate [11]. Therefore, predicting the potential associations between circRNAs and miRNAs can help biologists to understand complex pathogenesis of disease and further contribute to disease diagnoses.

With the continuous development of high-throughput sequencing technology, a large number of circRNAs have been discovered. Simultaneously, plenty of databases are developed to store circRNA-related information such as circBase [12], circR2Disease [13], circRNADisease [14], and circ2Disease [15]. circBase is an online database that provides users with a variety of basic circRNA information such as circRNA ID, sequence, gene description, and location [12]. circR2Disease is a public database that stores experimentally verified circRNA-related disease information. The database contains 793 circRNA-disease associations including 661 circRNAs and 100 diseases [13]. The circRNADisease database contains 354 circRNA-disease interactions, 330 circRNAs, and 48 diseases [14]. Similar to circR2Disease and circRNADisease, the circ2Disease database is used to store a vitro-proved circRNA-disease association database through which users can obtain circRNA-disease associations and the associations between miRNAs and its targets [15]. These databases enable users to identify potential associations between circRNA and miRNA by using computational methods.

Compared with traditional biological experiment methods, the circRNA-miRNA association prediction based on computational methods can maintain high accuracy and be less time-consuming. Therefore, more and more attention has been paid to circRNA-miRNA association prediction based on computational methods. At present, a large number of computational prediction models have been applied in many fields of biology, for example, predicting associations between diseases and genes, miRNA-disease associations [16, 17], circRNA-disease associations [18, 19], lncRNA-disease associations [20, 21], protein function [22, 23], drug-target interactions [24, 25], and lncRNA-miRNA associations [26, 27]. Compared with other fields, there are few prediction models based on the computational method in the circRNA-miRNA association prediction. Therefore, it is urgent to develop an effective computational method to infer circRNA-miRNA associations.

For the above purposes, in this study, we propose a new computational algorithm based on network embedding, NECMA, to predict circRNA-miRNA association. In our method, the circRNA-miRNA network is constructed based on experimental verified circRNA-miRNA associations. Then, based on the circRNA-miRNA associations, the GIP kernel similarities of circRNA and miRNA are calculated by using the Gaussian interaction profile kernel similarity, respectively. Furthermore, the circRNA GIP kernel similarity network, miRNA GIP kernel similarity network, and circRNA-miRNA association network are integrated to construct the circRNA-miRNA heterogeneous network. In addition, the network embedding model is employed to learn the features of circRNA and miRNA based on the circRNA-miRNA heterogeneous network, respectively. Finally, the weighted neighborhood regularized logistic matrix factorization and inner product are combined to predict potential circRNA-miRNA associations. The ten-fold cross-validation is used to evaluate the performance of our method. The experimental results show that NECMA achieves better performance than other state-of-the-art methods. In addition, the case study shows that NECMA could effectively infer potential circRNA-miRNA associations which are confirmed by the latest literature.

2. Related Work

Numerous experiments have shown that circRNA and miRNA have a close association with diseases. The current circRNA-disease association prediction algorithms are divided into the following categories. (1) Network-based circRNA-disease association prediction method: Fan et al. [28] used known circRNA-disease associations, circRNA expression profile similarities, and disease phenotype similarities to construct the circRNA-disease heterogeneous network and then used KATZ to predict potential associations between circRNAs and diseases. Li et al. [29] integrated known circRNA-disease associations, circRNA functional similarities, and disease semantic similarities and utilized network-consistent projections to identify potential circRNA-disease associations. Zhao et al. [30] developed an ensemble learning algorithm to predict the potential association between circRNA and diseases. In this method, the circRNA-disease heterogeneous network is constructed from known circRNA-disease association network, circRNA similarity network, and disease similarity network and circRNA-disease association is predicted by using KATZ and bipartite network projections. Lei and Bian [31] used random walk with restart and KNN algorithms to identify potential associations between circRNAs and diseases based on known circRNA-disease associations, circRNA similarities, and disease similarities. Li et al. [32] predicted potential circRNA-disease associations based on known circRNA-disease association networks, circRNA similarity, and disease similarity by using inductive matrix completion. Wei and Liu [18] reconstructed the circRNA-disease association network using circRNA similarities and disease similarities and then used nonnegative matrix factorization to predict potential associations. (2) Machine learning-based circRNA-disease association prediction method: Lei and Fang [33] fused circRNA expression profile similarity network, circRNA sequence similarity network, and circRNA functional annotation similarity network to construct the circRNA similarity network. The disease similarity network is constructed by integrating the disease functional similarity network and the disease semantic similarity network. Finally, based on the known circRNA-disease association network, circRNA similarity network, and disease similarity network, the potential feature of circRNA and diseases were extracted, respectively, and then the gradient boosting decision tree algorithm was used to predict the potential circRNA-disease associations.

Similarly, miRNA-disease association prediction algorithms can also be classified into the similar categories. Peng

et al. [34] developed a miRNA-disease association prediction model (ThrWRDE) that performs a restart random walk algorithm on a variety of miRNA-related biological data and then integrates the results obtained from multiple restart random walk models. You et al. [35] integrated known miRNA-disease associations, miRNA similarity, and disease similarity to construct a miRNA-disease heterogeneous network and then used a depth-first search algorithm to count the path between miRNA and diseases. Finally, the different pathways between miRNA and diseases are integrated to obtain the predicted association score between miRNA and diseases. Chen et al. [36] constructed miRNA similarity and disease similarity by integrating miRNA functional similarity and miRNA Gaussian interaction profile kernel similarity and disease Gaussian interaction profile kernel similarity and disease semantic similarity, respectively. Finally, the inductive matrix completion is used to obtain the final predicted miRNA-disease association. Chen et al. [37] extracted the potential representations of miRNA and disease, respectively, using a stacked autoencoder and then obtained the predicted score of miRNA-disease association by using support vector machine (SVM).

3. Materials and Methods

3.1. Materials. The circRNA-miRNA associations are downloaded from circR2Cancer database (http://www.biobdlab.cn:8000/). The circR2Cancer database [38] is a manually curated database which contains not only circRNA-cancer association data but also circRNA-miRNA association data and miRNA-cancer association data. After removing redundant data, 130 circRNAs, 412 miRNAs, and 477 associations are extracted in final. Furthermore, the adjacent matrix $CM_{m\times n}$ is constructed to represent circRNA-miRNA association, where m represents the number of circRNAs and n represents the number of miRNAs. The value of element $CM(i, j)$ is equal to 1 when circRNA i is related to miRNA j, otherwise 0.

3.2. circRNA and miRNA Similarity Calculation. In this study, the Gaussian interaction profile (GIP) kernel similarity is used to calculate similarities of circRNA and miRNA. Based on the assumption that circRNAs with similar functions are often associated with similar miRNAs, circRNA GIP kernel similarity and miRNA GIP kernel similarity are calculated based on the circRNA-miRNA interaction network, respectively. For pairwise circRNAs c_i and c_j, the GIP kernel similarity $CS(c_i, c_j)$ between circRNAs c_i and c_j is defined as follows:

$$CS\left(c_i, c_j\right) = \exp\left(-\gamma_c \left\|CM\left(c_i, :\right) - CM\left(c_j, :\right)\right\|^2\right), \quad (1)$$

where $CM(c_i, :)$ represents the c_i row in the matrix CM and γ_c represents the kernel bandwidth, which is defined as follows:

$$\gamma_c = \frac{1}{\left(1/n_c \sum_{i=1}^{n_c} \left\|CM\left(c_i, :\right)\right\|^2\right)}, \quad (2)$$

where n_c represents the number of rows in matrix CM.

Similarly, the miRNA GIP kernel similarity $MS(m_i, m_j)$ between miRNA m_i and miRNA m_j is defined as follows:

$$MS\left(m_i, m_j\right) = \exp\left(-\gamma_m \left\|CM\left(:, m_i\right) - CM\left(:, m_i\right)\right\|^2\right), \quad (3)$$

where $CM(:, m_i)$ represents m_i column of matrix CM and γ_m represents the kernel bandwidth, which is defined as follows:

$$\gamma_m = \frac{1}{\left(1/n_m \sum_{i=1}^{n_m} \left\|CM\left(:, m_i\right)\right\|^2\right)}, \quad (4)$$

where n_m represents the number of columns in the interaction matrix CM.

3.3. Construction of Heterogeneous Network. The heterogeneous network used for circRNA-miRNA association prediction is composed of three subnetworks including circRNA-miRNA interaction network, circRNA GIP kernel similarity network, and miRNA GIP kernel similarity network. Based on the above three subnetworks, the heterogeneous network H is constructed as follows:

$$H = \begin{bmatrix} CS & CM \\ CM^T & MS \end{bmatrix}, \quad (5)$$

where CM represents the circRNA-miRNA interaction network, CM^T represents the transpose of the circRNA-miRNA interaction network, CS represents the circRNA GIP kernel similarity network, and MS represents the miRNA GIP kernel similarity network.

3.4. The Feature Extraction Based on NetMF. After obtaining the heterogeneous network H, the network embedding as matrix factorization (NetMF) algorithm [39] is used to extract the potential features of circRNA and miRNA on the heterogeneous network, respectively. NetMF is a matrix factorization framework based on the original DeepWalk algorithm. To be specific, the NetMF model is the matrix factorization form of DeepWalk algorithm derived from the implicit decomposition model of the skip-gram with negative-sampling model (SGNS) [40, 41]. It can reduce the noise information in the matrix H and improve the performance of the prediction model. The NetMF model is defined as the probability distribution of truncated random walk, which is calculated as

$$\text{Pro} = D^{-1} \cdot H, \quad (6)$$

where D denotes a diagonal matrix and the elements in D represent the generalized degree of nodes in the matrix H.

Then, we conducted w times random walk on the heterogeneous network H according to the probability

distribution calculated before. It is used to sample the nodes in the heterogeneous network to obtain the transfer matrix trans_H which is defined as follows:

$$\text{trans\_}H = \sum_{m=1}^{w} \text{Pro}. \tag{7}$$

After obtaining the transition matrix trans_H, the DeepWalk matrix is obtained as

$$H_{\text{DeepWalk}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} H(i,j)}{bw} \cdot \text{trans\_}H \cdot D^{-1}, \tag{8}$$

where n represents the dimension of heterogeneous network H and b represents the number of negative samples.

Since the density of H_{DeepWalk} increases the time complexity of subsequent calculations, the approximate matrix H'_{DeepWalk} is defined as follows:

$$H'_{\text{DeepWalk}} = \max\left(H_{\text{DeepWalk}}, 1\right). \tag{9}$$

After obtaining the matrix H_{DeepWalk} of the circRNA-miRNA heterogeneous network H, the low-dimensional space feature vectors of circRNA and miRNA are obtained by using the singular value decomposition (SVD) model [42–44], which is defined as follows:

$$U_d \sum_d V_d^T = \log H'_{\text{DeepWalk}}, \tag{10}$$

where d represents the dimension of a low-dimensional space.

Finally, the eigenmatrix eigen_matrix is calculated:

$$\text{eigen\_matrix} = U_d \sum_d. \tag{11}$$

The dimension of eigen_matrix is $m + n$ and d. This matrix is composed of circRNA feature vectors u and miRNA feature vectors v, in which the dimensions are m and n, respectively.

3.5. circRNA-miRNA Association Prediction. The potential eigenvectors of circRNA and miRNA are obtained by NetMF on heterogeneous network H. Then, the weighted neighborhood regularized logistic matrix factorization [45] and inner product are utilized to reconstruct the circRNA-miRNA association matrix.

The weighted neighborhood regularized logistic matrix factorization is defined as follows:

$$\text{matrix\_pre}(i,j) = \frac{e^{\left((\alpha)u_i(1-\alpha)v_j^T\right)}}{1 + e^{\left((\alpha)u_i(1-\alpha)v_j^T\right)}}. \tag{12}$$

The inner product is defined as follows:

$$\text{inner\_pre}(i,j) = u_i v_j^T, \tag{13}$$

where u_i represents the feature vector of circRNA i and v_j represents the feature vector of miRNA j. α represents the weight coefficient to balance the influence of two feature vectors on the reconstructed matrix.

Finally, the score of circRNA-miRNA association matrix is defined as follows:

$$\text{Pre}(i,j) = \max(\text{matrix\_pre}(i,j), \text{inner\_pre}(i,j)), \tag{14}$$

where $\text{Pre}(i,j)$ denotes the predicted score between circRNA i and miRNA j.

The flowchart of NECMA is shown in Figure 1. It mainly contains the following steps: first, the Gaussian interaction profile kernel similarity is utilized to calculate circRNA similarity and miRNA similarity based on the known circRNA-miRNA associations, respectively. Then, the heterogeneous network H is constructed based on the circRNA-miRNA association network, miRNA similarity network, and circRNA similarity network. Furthermore, the NetMF is used to extract the low-dimensional features of circRNA and miRNA on heterogeneous network H, respectively. Finally, the weighted neighborhood regularized logistic matrix factorization and the inner product are utilized to reconstruct the circRNA-miRNA association matrix based on the circRNA feature vector and miRNA feature vector (Algorithm 1).

4. Result

4.1. Ten-Fold Cross-Validation. In order to evaluate the performance of NECMA, we conduct the ten-fold cross-validation in the experiment. In the ten-fold cross-validation, the known circRNA-miRNA associations are randomly divided into ten subsets. Then, in each round of cross-validation experiment, one set is selected as the test samples and the other nine sets are treated as the training samples which are used in model training. The final score of circRNA-miRNA association is predicted by using the model. The higher the score of the association, the higher the probability of circRNA-miRNA interaction. Then, we rearranged the score of circRNA-miRNA association in descending order. Furthermore, the true positive rate (TPR) and false positive rate (FPR) are calculated by altering the threshold. The TPR and FPR are defined as follows:

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{TN}}, \end{aligned} \tag{15}$$

where TP and FP represent true positive and false positive, respectively, and TN and FN represent true negative and false negative, respectively. Finally, the receiver operating characteristics (ROC) curve is plotted based on TPR and FPR, and the area under ROC curve (AUROC) is calculated to evaluate the predictive power of the model. The higher the AUROC value, the better the performance of the model.

Similarly, the area under precision-recall (AUPR) curve based on precision and recall is used to evaluate the performance of prediction model. The precision and recall are defined as follows:

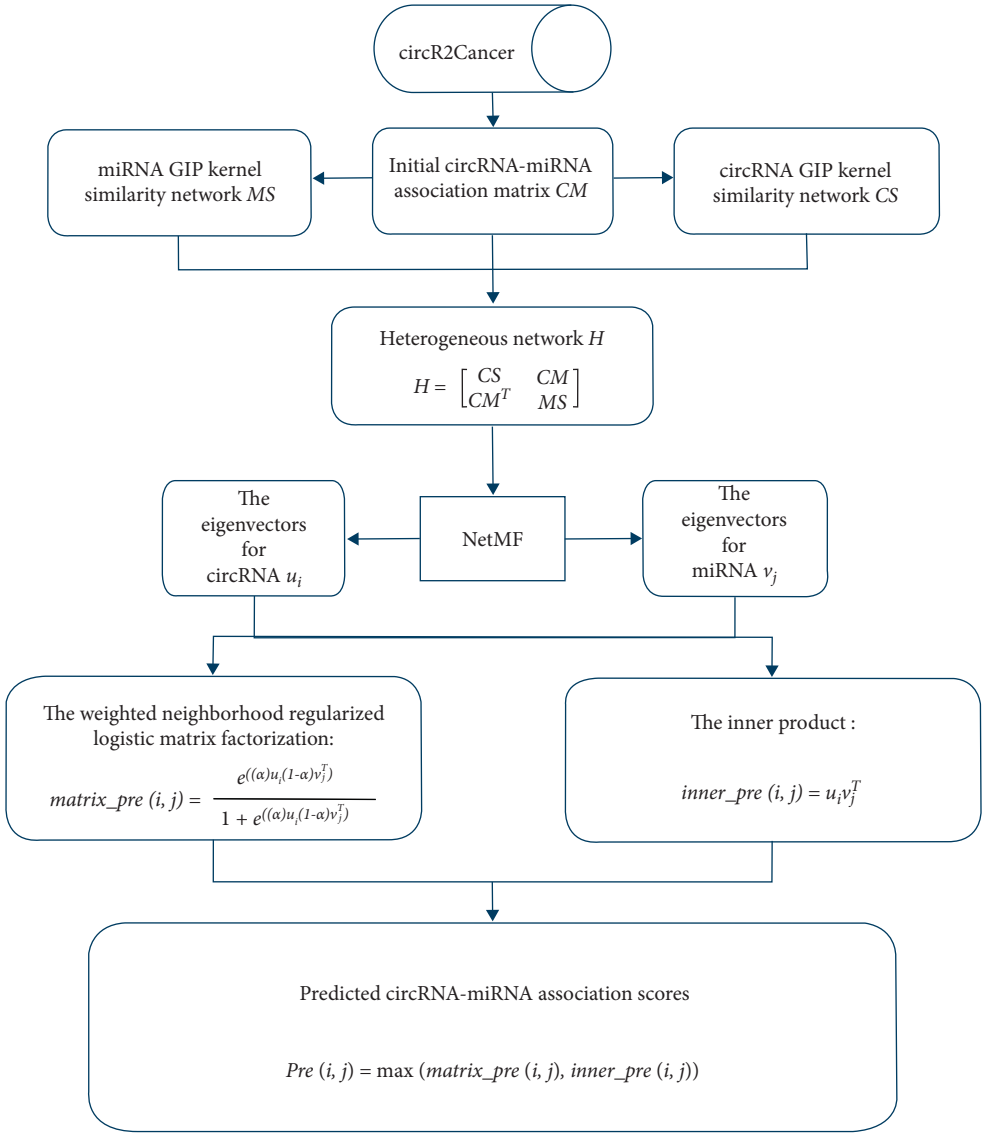


Figure 1: The flowchart of NECMA.

Input: circRNA-miRNA association matrix CM
Output: Predicted association matrix Pre
(1) Calculate the circRNA GIP kernel similarity CS based on known circRNA-miRNA associations
(2) Calculate the miRNA GIP kernel similarity MS based on known circRNA-miRNA associations
(3) Construct a heterogeneous network H based on CM, CS, and MS
(4) Calculate generalized degree matrix D based on heterogeneous network H
(5) Calculate the probability of truncated random walk Pro = $D^{-1} \cdot H$
(6) Calculate the transition matrix trans_H = $\sum_{m=1}^{w}$ Pro
(7) Calculate $H_{\text{DeepWalk}} = (\sum_{i=1}^{n} \sum_{j=1}^{n} H(i,j)/bw) \cdot \text{trans\_}H \cdot D^{-1}$
(8) Calculate $H'_{\text{DeepWalk}} = \max(H_{\text{DeepWalk}}, 1)$
(9) Dimensionality reduction through SVD $U_d \sum_d V_d^T = \log H'_{\text{DeepWalk}}$
(10) Calculate circRNA-miRNA feature matrix eigen_matrix = $U_d \sum_d$
(11) Extract circRNA feature u and miRNA feature v based on eigen_matrix
(12) Calculate matrix_pre$(i, j) = e^{((\alpha)u_i(1-\alpha)v_j^T)}/1 + e^{((\alpha)u_i(1-\alpha)v_j^T)}$
(13) Calculate inner_pre$(i, j) = u_i v_j^T$
(14) **Return:** Predicted association matrix: Pre$(i, j) = \max(\text{matrix_pre}(i, j), \text{inner_pre}(i, j))$

Algorithm 1: NECMA.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$
$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{16}$$

where precision represents the proportion of positive examples in the predicted results to the actual positive examples and recall represents the proportion of all true positive cases divided into positive cases, which measures the classifier's ability to recognize positive cases.

In addition, in order to demonstrate the superiority of NECMA in predicting the potential association of circRNA-miRNA. We compare NECMA with three state-of-the-art algorithms including RWRLncD [46], NCPLDA [47], and LRLSLDA [48]. Figures 2 and 3 show the AUROC and AUPR values obtained by different prediction models in ten-fold cross-validation, respectively. The results of ten-fold cross-validation show that the AUROC and AUPR of the NECMA are better than other three prediction algorithms. It can be found from Figure 2 that the AUROC value of NECMA is 0.8264 which is higher than RWRLncD (0.5243), NCPLDA (0.6985), and LRLSLDA (0.7661). Simultaneously, it can be observed from Figure 3 that the AUPR value of NECMA is 0.0048 which is higher than RWRLncD (0.0016), NCPLDA (0.0011), and LRLSLDA (0.0026). The overall results of ten-fold cross-validation are shown in Table 1. It can be concluded that NECMA is an effective method in identifying association between circRNA and miRNA.

4.2. Effect of Parameters. There are three parameters contained in the NetMF model (context window w, negative sampling number b, and embedded dimension d). According to the previous study [39], both the context window w and the negative sampling number b are set to 1. To test the effect of embedded dimension d, we set embedded dimension d ranging from 8 to 128. The result is shown in Figure 4. It can be found that the AUROC value of NECMA is the highest when the embedded dimension d of the NetMF is set to 8. In addition, we also test the effect of parameter weight coefficient α in neighborhood regularization logistic matrix factorization which is used to balance the influence of two eigenvectors in the process of matrix reconstruction. The parameter α ranges from 0.1 to 0.9 with 0.1 increasing in each time. The influence of parameter α on the prediction performance is shown in Figure 5. It can be observed that the AUROC value obtained is the highest when $\alpha = 0.6$.

4.3. Case Study. To further illustrate the ability of NECMA to predict potential circRNA-miRNA associations, we conduct a case study on miR-130a-3p. We select the top 10 circRNAs predicted by NECMA and prove these associations by manually retrieving related databases and literature.

The numerous experiments have shown that miR-130a-3p is associated with proliferation and migration of many cancer cells [49]. For example, miR-130a-3p can regulate its target Smad4 to inhibit migration and invasion of gemcitabine-resistant (GR) hepatocellular carcinoma (HCC) cells [50]. Therefore, the correct prediction of the circRNAs associated with miR-130a-3p is useful for understanding complex disease mechanisms. The top 10 predicted circRNAs of miR-130a-3p are shown in Table 2. The results showed that nine circRNAs (hsa_circ_0068942, hsa_circ_0089378, hsa_circ_0083357, hsa_circ_0006323, hsa_circ_0032970, hsa_circ_0051172, hsa_circ_0054537, hsa_circ_0057576, and hsa_circ_0082824) have been confirmed in the literature. It has been confirmed that the moderately upregulated hsa_circ_0068942 ranked at top 1 can serve as miR-130a-3p sponge and the disease marker for coronary artery disease (CAD) [51]. It has been demonstrated that hsa_circ_0089378 ranked at top 2 can act as

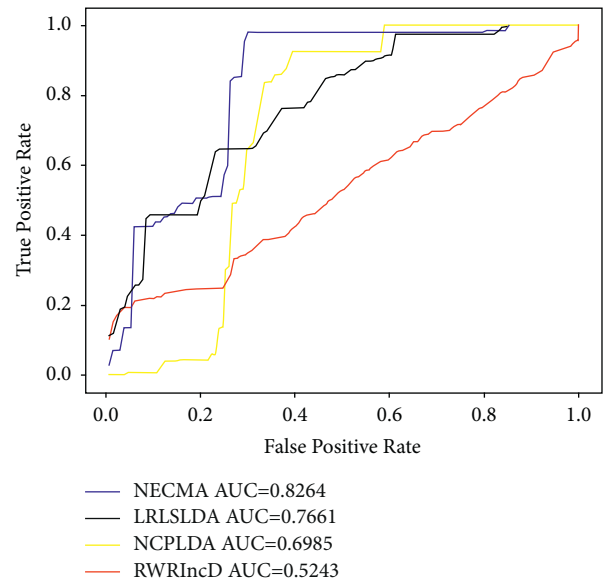


Figure 2: The AUROC values of the NECMA model and the other three algorithms on the ten-fold cross-validation.

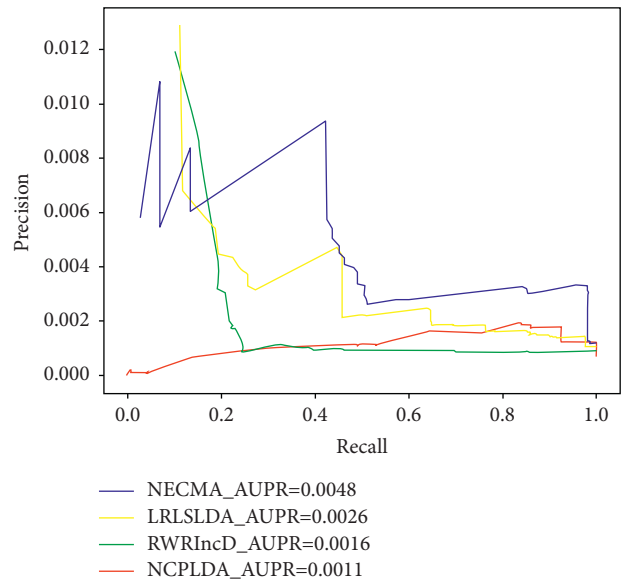


Figure 3: The AUPR values of the NECMA model and the other three algorithms on the ten-fold cross-validation.

Table 1: Performance comparison of NECMA with RWRLncD, LRLSLDA, and NCPLDA.

Methods	AUC	AUPR	Precision	Recall	Accuracy	*F1*-score	Specificity
RWRLncD	0.5243	0.0016	0.0013	0.5321	0.4922	0.0026	0.4922
LRLSLDA	0.7661	0.0026	0.0019	0.7720	0.4926	0.0039	0.4924
NCPLDA	0.6958	0.0011	0.0009	0.7005	0.4981	0.0018	0.4979
NECMA	**0.8264**	**0.0048**	**0.0026**	**0.8277**	**0.4993**	**0.0053**	**0.4989**

The bold values indicate best performance.

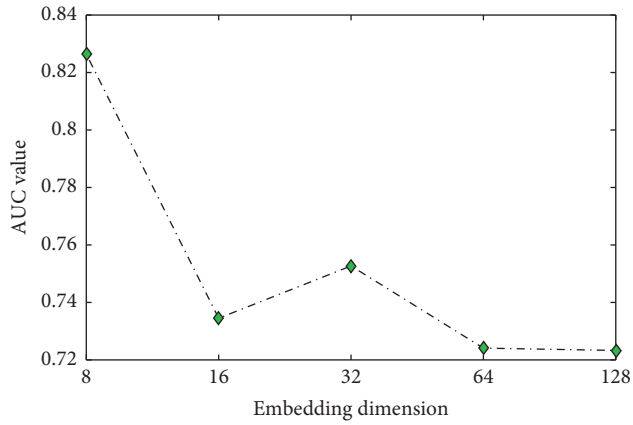


Figure 4: The impact of embedded dimension d.

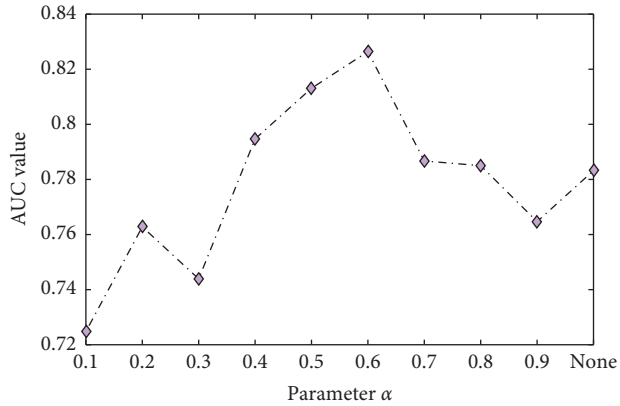


Figure 5: The influence of parameter α.

Table 2: Top ten candidate circRNAs for miR-130a-3p.

Rank	circRNA	Evidence
1	hsa_circ_0068942	PMID: 30159442
2	hsa_circ_0089378	PMID: 31119072
3	hsa_circ_0083357	PMID: 30368217
4	hsa_circ_0006323	PMID: 29730164
5	hsa_circ_0032970	PMID: 28947970
6	hsa_circ_0051172	PMID: 29182528
7	hsa_circ_0054537	PMID: 28947970
8	hsa_circ_0057576	PMID: 28947970
9	hsa_circ_0082824	PMID: 29182528
10	hsa_circ_0005986	Uknown

the sponge for miR-130a-3p to affect its target mRNA expression in coronary artery disease [52]. It has been demonstrated that hsa_circ_0083357 ranked at top 3 can play an important role in coronary artery disease through miR-130a-3p-mediated circRNA-mRNA-competitive endogenous RNA (ceRNA) networks [53]. It has been found that hsa_circ_0006323 ranked at top 4 can inhibit the expression of miR-130a-3p in coronary artery disease cells

[54]. It has been demonstrated that hsa_circ_0032970 ranked at top 5 can bind to miR-130a-3p binding sites in coronary artery disease cells [55]. It has been demonstrated that hsa_circ_0051172 ranked at top 6 can regulate the expression of TRPM3 by targeting miR-130a-3p in coronary artery disease [56]. In addition, it has been confirmed that differential expression of hsa_circ_00 54537 ranked at top 7 and hsa_circ_0057576 ranked at top 8 can not only inhibit miR-130a-3p but also lead to upregulation of TRPM3 [55]. It has been discovered that hsa_circ_0082824 ranked at top 9 can promote the expression of TRPM3 in target cells in coronary artery disease by inhibiting miR-130a-3p [54].

5. Conclusion

Accumulating experiments have shown that predicting associations between circRNAs and miRNAs not only helps to understand complex disease mechanisms but also contributes to prevent and diagnose diseases [57]. In this study, we propose a computational method, NECMA, to infer circRNA-miRNA associations. In this model, we first construct the circRNA-miRNA association matrix based on known circRNA-miRNA associations. Then, the Gaussian interaction profile kernel similarity is used to calculate circRNA similarity and miRNA similarity based on known circRNA-miRNA associations, respectively. Furthermore, the heterogeneous network is constructed based on three subnetworks (circRNA-miRNA association network, circRNA similarity network, and miRNA similarity network). In addition, the NetMF is employed to extract the subspace features of circRNA and miRNA from the heterogeneous network, respectively. Finally, the scores of circRNA-miRNA associations are predicted by using weighted neighborhood regularized logistic matrix factorization and inner product. In order to show the performance of NECMA, we compare NECMA with three state-of-the-art methods (RWRLncD, NCPLDA, and LRLSLDA) in terms of ten-fold cross-validation. The experimental results show that the NECMA achieves a higher AUROC value (0.8264) than the other three prediction models. In addition, it is demonstrated that NECMA could correctly identify potential associations between circRNA and miRNAs by constructing a case study on miR-130a-3p.

Although the NECMA model can effectively predict the potential circRNA-miRNA association, there are still many limitations. First, the NECMA model mainly relies on known circRNA-miRNA association data and the imbalance of positive and negative samples will greatly affect the prediction accuracy of the model. Second, the setting of parameters will also affect the prediction results of the model. In addition, the integration of various circRNA and miRNA information can further improve the predictive power of the model [58–60]. Moreover, the NECMA model cannot predict new circRNA-miRNA without any known association. Therefore, we will integrate more biological data of circRNA and miRNA in the future, which will make it more reliable [61–63].

Data Availability

The underlying data supporting the results of our study can be found at http://www.biobdlab.cn:8000/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Nos. 62072124, 61963004, and 61972185), the Natural Science Foundation of Guangxi (Nos. 2021GXNSFAA075041 and 2018GXNSFBA281193), the Science and Technology Base and Talent Special Project of Guangxi (No. AD20159044), the Natural Science Foundation of Yunnan Province of China (No. 2019FA024), the Hunan Provincial Science and Technology Program (No. 2018WK4001), and the Scientific Research Foundation of Hunan Provincial Education Department (No. 18B469).

References

[1] S. Qu, X. Yang, X. Li et al., "Circular RNA: a new star of noncoding RNAs," *Cancer Letters*, vol. 365, no. 2, pp. 141–148, 2015.

[2] L. H. Jiang, D. W. Sun, J. C. Hou, H.-d. Zhang, and Z. L. Ji, "CircRNA: a novel type of biomarker for cancer," *Breast Cancer*, vol. 25, no. 1, pp. 1–7, 2018.

[3] M. Hanan, H. Soreq, and S. Kadener, "CircRNAs in the brain," *RNA Biology*, vol. 14, no. 8, pp. 1028–1034, 2017.

[4] L.-L. Chen, "The biogenesis and emerging roles of circular RNAs," *Nature Reviews Molecular Cell Biology*, vol. 17, no. 4, pp. 205–211, 2016.

[5] A. A. Nair, N. Niu, X. Tang et al., "Circular RNAs and their associations with breast cancer subtypes," *Oncotarget*, vol. 7, no. 49, pp. 80967–80979, 2016.

[6] S. Memczak, M. Jens, A. Elefsinioti et al., "Circular RNAs are a large class of animal RNAs with regulatory potency," *Nature*, vol. 495, no. 7441, pp. 333–338, 2013.

[7] G. Huang, S. Li, N. Yang, Y. Zou, D. Zheng, and T. Xiao, "Recent progress in circular RNAs in human cancers," *Cancer Letters*, vol. 404, pp. 8–18, 2017.

[8] W. Weng, Q. Wei, S. Toden et al., "Circular RNA ciRS-7-A promising prognostic biomarker and a potential therapeutic target in colorectal cancer," *Clinical Cancer Research*, vol. 23, no. 14, pp. 3918–3928, 2017.

[9] G. Chen, Y. Shi, M. Liu, and J. Sun, "circHIPK3 regulates cell proliferation and migration by sponging miR-124 and regulating AQP3 expression in hepatocellular carcinoma," *Cell Death & Disease*, vol. 9, no. 2, pp. 175–213, 2018.

[10] W. Lan, J. Wang, M. Li, W. Peng, and F. Wu, "Computational approaches for prioritizing candidate disease genes based on PPI networks," *Tsinghua Science and Technology*, vol. 20, no. 5, pp. 500–512, 2015.

[11] J. Zhang, H. Guo, H. Zhang et al., "Putative tumor suppressor miR-145 inhibits colon cancer cell growth by targeting oncogene friend leukemia virus integration 1 gene," *Cancer*, vol. 117, no. 1, pp. 86–95, 2011.

[12] P. Glažar, P. Papavasileiou, and N. Rajewsky, "circBase: a database for circular RNAs," *RNA*, vol. 20, no. 11, pp. 1666–1670, 2014.

[13] C. Fan, X. Lei, Z. Fang, Q. Jiang, and F. X. Wu, "CircR2-Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases," *Database*, vol. 2018, pp. 1–6, 2018.

[14] Z. Zhao, K. Wang, F. Wu et al., "circRNA disease: a manually curated database of experimentally supported circRNA-disease associations," *Cell Death & Disease*, vol. 9, no. 5, pp. 475–482, 2018.

[15] D. Yao, L. Zhang, M. Zheng, X. Sun, Y. Lu, and P. Liu, "Circ2Disease: A manually curated database of experimentally validated circRNAs in human disease," *Scientific Reports*, vol. 8, no. 1, pp. 1–6, 2018.

[16] W. Lan, J. Wang, M. Li, J. Liu, F.-X. Wu, and Y. Pan, "Predicting microRNA-disease associations based on improved microRNA and disease similarities," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 6, pp. 1774–1782, 2018.

[17] W. Peng, W. Lan, J. Zhong, J. Wang, and Y. Pan, "A novel method of predicting microRNA-disease associations based on microRNA, disease, gene and environment factor networks," *Methods*, vol. 124, pp. 69–77, 2017.

[18] H. Wei and B. Liu, "iCircDA-M. F.: iCircDA-MF: identification of circRNA-disease associations based on matrix factorization," *Briefings in Bioinformatics*, vol. 21, no. 4, pp. 1356–1367, 2020.

[19] C. Yan, J. Wang, and F. X. Wu, "DWNN-RLS.: regularized least squares method for predicting circRNA-disease associations," *BMC Bioinformatics*, vol. 19, no. 19, Article ID 520, 2018.

[20] W. Lan, M. Li, K. Zhao et al., "LDAP: a web server for lncRNA-disease association prediction," *Bioinformatics*, vol. 33, no. 3, pp. 458–460, 2017.

[21] X. Wu, W. Lan, Q. Chen, Y. Dong, J. Liu, and W. Peng, "Inferring LncRNA-disease associations based on graph autoencoder matrix completion," *Computational Biology and Chemistry*, vol. 87, Article ID 107282, 2020.

[22] R. Cao and J. Cheng, "Integrated protein function prediction by mining function associations, sequences, and protein-protein and gene-gene interaction networks," *Methods*, vol. 93, pp. 84–91, 2016.

[23] W. Peng, J. Du, L. Li, W. Dai, and W. Lan, "Predicting protein functions by using non-negative matrix factorisation with multi-networks co-regularisation," *International Journal of Data Mining and Bioinformatics*, vol. 23, no. 4, pp. 318–342, 2020.

[24] W. Lan, J. Wang, M. Li et al., "Predicting drug-target interaction using positive-unlabeled learning," *Neurocomputing*, vol. 206, pp. 50–57, 2016.

[25] C. Yan, J. Wang, W. Lan, F. X. Wu, and Y. Pan, "SDTRLS: predicting drug-target interactions for complex diseases based on chemical substructures," *Complexity*, vol. 2017, Article ID 2713280, 10 pages, 2017.

[26] H. Liu, G. Ren, H. Chen, Q. Liu, Y. Yang, and Q. Zhao, "Predicting lncRNA-miRNA interactions based on logistic matrix factorization with neighborhood regularized," *Knowledge-Based Systems*, vol. 191, Article ID 105261, 2020.

[27] W. Lan, L. Huang, D. Lai, and Q. Chen, "Identifying interactions between long noncoding RNAs and diseases based on computational methods," *Methods in Molecular Biology*, vol. 1754, pp. 205–221, 2018.

[28] C. Fan, X. Lei, and F.-X. Wu, "Prediction of CircRNA-disease associations using KATZ model based on heterogeneous networks," *International Journal of Biological Sciences*, vol. 14, no. 14, pp. 1950–1959, 2018.

[29] G. Li, Y. Yue, C. Liang, Q. Xiao, P. Ding, and J. Luo, "NCPCDA: network consistency projection for circRNA-disease association prediction," *RSC Advances*, vol. 9, no. 57, pp. 33222–33228, 2019.

[30] Q. Zhao, Y. Yang, G. Ren, E. Ge, and C. Fan, "Integrating bipartite network projection and KATZ measure to identify novel CircRNA-disease associations," *IEEE Transactions on Nanobioscience*, vol. 18, no. 4, pp. 578–584, 2019.

[31] X. Lei and C. Bian, "Integrating random walk with restart and k-nearest neighbor to identify novel circRNA-disease association," *Scientific Reports*, vol. 10, no. 1, pp. 1943–1949, 2020.

[32] M. Li, M. Liu, Y. Bin, and J. Xia, "Prediction of circRNA-disease associations based on inductive matrix completion," *BMC Medical Genomics*, vol. 13, no. 5, pp. 42–13, 2020.

[33] X. Lei and Z. Fang, "GBDTCDA: predicting circRNA-disease associations based on gradient boosting decision tree with multiple biological data fusion," *International Journal of Biological Sciences*, vol. 15, no. 13, pp. 2911–2924, 2019.

[34] W. Peng, W. Lan, Z. Yu, J. Wang, and Y. Pan, "A framework for integrating multiple biological networks to predict MicroRNA-disease associations," *IEEE Transactions on Nanobioscience*, vol. 16, no. 2, pp. 100–107, 2016.

[35] Z.-H. You, Z.-A. Huang, Z. Zhu et al., "PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction," *PLoS Computational Biology*, vol. 13, no. 3, Article ID e1005455, 2017.

[36] X. Chen, L. Wang, J. Qu, N. N. Guan, and J. Q. Li, "Predicting miRNA–disease association based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 24, pp. 4256–4265, 2018.

[37] X. Chen, Y. Gong, D.-H. Zhang, Z.-H. You, and Z.-W. Li, "DRMDA: deep representations-based miRNA-disease association prediction," *Journal of Cellular and Molecular Medicine*, vol. 22, no. 1, pp. 472–485, 2018.

[38] W. Lan, M. Zhu, Q. Chen et al., "CircR2Cancer: a manually curated database of associations between circRNAs and cancers," *Database*, vol. 2020, 2020.

[39] J. Qiu, Y. Dong, H. Ma, J. Li, K. Wang, and J. Tang, "Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM'18)*, pp. 459–467, Marina Del Rey, CA, USA, February 2018.

[40] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[41] J. Wu, S. Pan, X. Zhu, and Z. Cai, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2014.

[42] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, vol. 14, no. 5, pp. 403–420, 1970.

[43] C. Y. Liu, C. Zhou, J. Wu, Y. Hu, and L. Guo, "Social recommendation with an essential preference space," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, New Orleans, LA, USA, Febraury 2018.

[44] L. Gao, J. Wu, Z. Qiao, C. Zhou, H. Yang, and Y. Hu, "Collaborative social group influence for event recommendation," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 1941–1944, Indianapolis, IA, USA, October 2016.

[45] Y. Liu, M. Wu, C. Miao, P. Zhao, and X. L. Li, "Neighborhood regularized logistic matrix factorization for drug-target interaction prediction," *PLoS Computational Biology*, vol. 12, no. 2, 2016.

[46] J. Sun, H. Shi, Z. Wang et al., "Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network," *Molecular BioSystems*, vol. 10, no. 8, pp. 2074–2081, 2014.

[47] G. Li, J. Luo, C. Liang, Q. Xiao, P. Ding, and Y. Zhang, "Prediction of LncRNA-disease associations based on network consistency projection," *IEEE Access*, vol. 7, pp. 58849–58856, 2019.

[48] X. Chen and G.-Y. Yan, "Novel human lncRNA-disease association inference based on lncRNA expression profiles," *Bioinformatics*, vol. 29, no. 20, pp. 2617–2624, 2013.

[49] X. Chen, B. Yue, C. Zhang et al., "MiR-130a-3p inhibits the viability, proliferation, invasion, and cell cycle, and promotes apoptosis of nasopharyngeal carcinoma cells by suppressing BACH2 expression," *Bioscience Reports*, vol. 37, no. 3, 2017.

[50] Y. Liu, Y. Li, R. Wang et al., "MiR-130a-3p regulates cell migration and invasion via inhibition of Smad4 in gemcitabine resistant hepatoma cells," *Journal of Experimental & Clinical Cancer Research: Climate Research*, vol. 35, no. 1, pp. 19–11, 2016.

[51] L. M. Holdt, A. Kohlmaier, and D. Teupser, "Molecular functions and specific roles of circRNAs in the cardiovascular system," *Non-Coding RNA Research*, vol. 3, no. 2, pp. 75–98, 2018.

[52] W. Wang, Y. Wang, H. Piao et al., "Circular RNAs as potential biomarkers and therapeutics for cardiovascular disease," *PeerJ*, vol. 7, Article ID e6831, 2019.

[53] Y. Bei, T. Yang, L. Wang et al., "Circular RNAs as potential theranostics in the cardiovascular system," *Molecular Therapy-Nucleic Acids*, vol. 13, pp. 407–418, 2018.

[54] K. Lei, H. Bai, Z. Wei et al., "The mechanism and function of circular RNAs in human diseases," *Experimental Cell Research*, vol. 368, no. 2, pp. 147–158, 2018.

[55] R.-Y. Pan, P. Liu, H.-T. Zhou et al., "Circular RNAs promote TRPM3 expression by inhibiting hsa-miR-130a-3p in coronary artery disease patients," *Oncotarget*, vol. 8, no. 36, pp. 60280–60290, 2017.

[56] S. Haque and L. Harries, "Circular RNAs (circRNAs) in health and disease," *Genes*, vol. 8, no. 12, Article ID 353, 2017.

[57] X. Lei, T. B. Mudiyanselage, Y. Zhang et al., "A comprehensive survey on computational methods of non-coding RNA and disease association prediction," *Briefings in Bioinformatics*, vol. 22, no. 4, Article ID bbaa350, 2020.

[58] W. Lan, D. Lai, Q. Chen et al., "LDICDL: LncRNA-disease association identification based on collaborative deep learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–9, 2020.

[59] W. Lan, Q. Chen, T. Li, C. Yuan, S. Mann, and B. Chen, "Identification of important positions within miRNAs by integrating sequential and structural features," *Current Protein & Peptide Science*, vol. 15, no. 6, pp. 591–597, 2014.

[60] Q. Chen, W. Lan, and J. Wang, "Mining featured patterns of MiRNA interaction based on sequence and structure similarity," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 2, pp. 415–422, 2013.

[61] R. Zheng, M. Li, X. Chen et al., "An ensemble method to reconstruct gene regulatory networks based on multivariate adaptive regression splines," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 347–354, 2021.

[62] J. Liu, D. Zeng, R. Guo, M. Lu, F. X. Wu, and J. Wang, "MMHGE: detecting mild cognitive impairment based on multi-atlas multi-view hybrid graph convolutional networks and ensemble learning," *Cluster Computing*, vol. 24, no. 3, pp. 103–113, 2020.

[63] J. Liu, M. Li, W. Lan, F. X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimer's disease using whole brain hierarchical network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 2, pp. 624–632, 2016.

Hindawi
Complexity
Volume 2021, Article ID 8859225, 14 pages
https://doi.org/10.1155/2021/8859225

Research Article

A Novel Emerging Topic Identification and Evolution Discovery Method on Time-Evolving and Heterogeneous Online Social Networks

Xiaoyan Xu,[1] **Wei Lv**,[1] **Beibei Zhang**,[2] **Shuaipeng Zhou**,[3] **Wei Wei**,[2] **and Yusen Li**[1]

[1]*School of Science, Xi'an Shiyou University, Xi'an 710056, China*
[2]*School of Computer and Engineering, Xi'an University of Technology, Xi'an 710048, China*
[3]*Aamaze Data Company, Xi'an, China*

Correspondence should be addressed to Beibei Zhang; beibeizhang115@hotmail.com

Received 3 September 2020; Revised 4 July 2021; Accepted 2 August 2021; Published 27 August 2021

Academic Editor: Jia Wu

With the fast development of web 2.0, information generation and propagation among online users become deeply interweaved. How to effectively and immediately discover the new emerging topic and further how to uncover its evolution law are still wide open and urgently needed by both research and practical fields. This paper proposed a novel early emerging topic detection and its evolution law identification framework based on dynamic community detection method on time-evolving and scalable heterogeneous social networks. The framework is composed of three major steps. Firstly, a time-evolving and scalable complex network denoted as KeyGraph is built up by deeply analyzing the text features of all kinds of data crawled from heterogeneous online social network platforms; secondly, a novel dynamic community detection method is proposed by which the new emerging topic is detected on the modeled time-evolving and scalable KeyGraph network; thirdly, a unified directional topic propagation network modeled by a great number of short texts including microblogs and news titles is set up, and the topic evolution law of the previously detected early emerging topic is identified by fully utilizing local network variations and modularity optimization of the "time-evolving" and directional topic propagation network. Our method is proved to yield preferable results on both a huge amount of computer-generated test data and a great amount of real online network data crawled from mainstream heterogeneous social networks.

1. Introduction

In recent years, with the fast development of web 2.0, social network sites such as Facebook, Sina microblog, and Twitter rise in a short time, a huge heterogeneous online social networks have gradually formed on which the functional role of online users is changing from the information consumers to both diffusers and generators [1]. Information of different online social networks propagates in a deeply mingled way. For example, information from news websites is reposted to sites like personal (micro) blogs that have specific following groups; BBS (bulletin board system) of which the information bares the broadcasting attribute is shared to the personal (micro) blogs. Activities above making the information production and propagation among various users become a huge data-heterogeneous, time-evolving, and scalable complex network. Thus, how to efficiently and timely identify and reveal the new emerging topic and even its evolution process (law) on this scalable, time-evolving, and heterogeneous online social network has become the hot research spot in topic detection and tracking fields lately.

It is widely known that a complex network bears the community structure, which represents a clustering of network nodes with densely interweaved edges within groups but spares connections between them [2]. Community structure not only reveals the coarse structure of the network but also plays an important role in the functioning of the network [3, 4]. For example, community in the social

network represents the real social groups composed of people having the same backgrounds or interests; community in collaboration network represents the related papers with the same research subject; community in the biology or circuit network represents nodes group with the same network function. Deng et al. [5] proposed a hot topic detection algorithm based on a community of networks. Lin and Guan-Zhong. [6] also found the forum hot topic using the community detection concept applied on the BBS network and validated the efficiency and consistency by manual calibration of the identified topic and the community. Some researchers [7, 8] discovered the epidemic spreading mechanism by using the community structure analysis of complex networks. Identification of community structure in the complex network usually can realize certain application-based purposes and thus, the community identification and topology evolution discovery have become the most significant focus in the complex network structure analysis field. Although community detection/identification in networks has been studied for many years, most existed approaches are designed simply for the static network [9, 10] and unified homogeneous network [11]. However, in the real-world and in this paper, the studied social network is time-evolving and heterogeneous due to the time-varying social communications and the time-dependent interactions of different social network platforms.

Although it has been concluded that the traditional static community detection methods could be applied to the time-evolving heterogeneous network by converting the time-evolving network into a series of static snapshots of sub-networks through rearranging the nodes and links belonging to the same time stamps, while in this process, the semantic relationships and dynamic properties of communities may have violently been damaged or even lost. Another concluded method is to identify network community not from scratch but by storing and using the historical results of execution of static community detection algorithm along with the network evolving process [12, 13]. However, a great amount of time and space computing costs are needed and the efficiency of the algorithm becomes slower with time flying. Lately, community discovery in the time-evolving and heterogeneous network has emerged as an outstanding challenge and has attracted much attention of researchers. Sun et al. [14] proposed a Dirichlet Process Mixture Model based algorithm to describe the community detection in a heterogeneous star-model network.

In summary, when applying existing community detection methods for time-evolving and heterogeneous networks, three main problems are usually encountered: (1) most existed community detection methods are proposed simply for the static and homogeneous network; (2) the semantic relationships and dynamic properties of communities are violently damaged and even bluntly lost due to the man-made segmentation of network; (3) a great amount of computing time and space cost is required by storing the historical community structure information as the initial input values.

Aiming to tackle these problems, we propose an emerging topic identification and evolution topology discovery framework based on a novel dynamic community detection method on the time-evolving and heterogeneous social network. Firstly, a unified short-text network is constructed by modeling heterogeneous short texts crawled from different online social networks into a network. It is denoted as KeyGraph based on the cooccurrence of keywords of the crawled short texts. Secondly, a novel dynamic community detection method with a well-known static community detection algorithm as the Louvain algorithm corely embedded is proposed and is applied on the KeyGraph, as a result of which, the new emerging topic in the form of newborn community is detected. Finally, the topic evolution topology is discovered by deep analysis of the community scale and nodes variation with time-evolving of the detected communities.

The rest of this paper is presented as follows. We briefly review the related research work in Section 2, our methodology is presented in Section 3, numerical results and evaluations are presented in Section 4, and finally, we conclude and discuss our future work in Section 5.

2. Related Work

Topic detection and tracking (TDT) is firstly put forward by DARPA in 1996. Its original objective is to automatically identify online public sentiments in the form of a topic from the network media stream and further to track the propagation and diffusion process of the previously identified/detected topic. Later, TDT has become the key technology in the fields of Internet public opinion/sentiment mining field. Classic TDT methods include the latent Dirichlet analysis (LDA) [15] and the probabilistic latent semantic analysis (PLSA) [16] methods. Their main concepts are that a topic is based on the probability distributions of sets of words such that the probability distribution of word cooccurrences among sets of words is maximized/optimized.

Community is another important property of a complex network except for the small-world and scale-free properties. It not only provides a coarse view of the network structure but also actually plays and represents certain functions of the network. Community describes the closeness among nodes which means that nodes are closely interrelated within groups, while nodes between different communities are loosely connected.

Sayyadi and Raschid [17] proposed a graph analytical method to detect and identify the topic; they proposed the KeyGraph algorithm to transform original texts data into a term graph based on properties of cooccurrence relations of texts data with each other. Furthermore, they utilized a community detection method to part the constructed KeyGraph network into community topology and they deemed each identified community as a detected topic. Nowadays, community detection is a fundamental technique of network structure analysis, many creative methods for discovering communities in a static and homogeneous network have been deployed in the past decades. It can be commonly divided into two classes: graph theory-based algorithm and sociology-based algorithm. Sociology-based algorithms can be generally divided into division and aggregation methods. The classical GN algorithm [2] belongs

to the division method, and its fundamental principle is to obtain network community by finding the edge with the highest score of betweenness and by removing it from the network. Newman proposed a fast aggregation algorithm [9], which has similar accuracy with GN, but the performance has been significantly improved. Blonde et al. [18] proposed the Louvain algorithm based on the modularity optimization method, which is a simple, efficient, and easy-to-implement method for finding community structure in a large-scale network. The method is actually a greedy optimization method that attempts to optimize the tag-indexed "modularity" of every possible partition of the target network.

However, the resolution limit problem is commonly encountered using the modularity-based community detection method in a static and homogeneous network. Here, the "static" mainly refers to both the time and the whole network structure staying static without variation with time flying by. The resolution limit problem means when the scale of the network is large enough, a small community in a large network cannot be properly and efficiently detected, which results in the overlapping community phenomena. This phenomenon is called the resolution limit problem in a large-scale network when using the modularity method for network community detection purposes [19, 20]. These community detection methods, however, lack the capability of dealing with time-evolving and cannot be directly used for heterogeneous networks. As our problem of identifying and revealing the emerging topic and its evolution topology on the large-scale, time-varying, and heterogeneous online social network, the previous and classic community detection methods face great challenges.

2.1. Dynamic Community Detection of Homogeneous Social Network. Tracking the evolution topology of detected community need to take the dynamic property of the time-evolving network into consideration. A commonly utilized and concluded framework [21–24] is to apply the static community detection algorithms for each static snapshots subnetwork composed of nodes and edges with the same timestamp of the time-evolving network and then to generate the evolution of community by computing the community closeness between two adjacent static snapshots subnetworks. Toyoda and Kitsuregawa [25] firstly selected these web pages with high focus numbers (thumb-up number) as seed web pages and finally obtained the community including the seed web pages by utilizing the page closeness calculation algorithm with hyperlink-induced topic search as core calculation framework. Palla et al. [26] obtained the community topology of one snapshot using the clique percolation clustering method and evaluated its usefulness by applying it to the scientists' cooperation network and the telecommunication users network. Chakrabarti et al. [27] proposed an evolution clustering model with k-means and hierarchy clustering method to identify the community evolution law in the process of dynamic community detection.

Another policy to track community evolution in the time-evolving network is by integrating the optimization of both modularity and the structure of local network variations into a multiobjective optimization problem. Its main concept is that treating the community topology of the previous timestamp $t - \Delta t$ as baseline network, at the present time epoch t, the variations of the network during the time range Δt, i.e., $[t - \Delta t, t]$ are the main focus other than the whole network of the present time being t. By detecting the community topology of the variation part of the network during the time range Δt while the other part of the network remains unchanged to improve the whole efficiency of the community detection algorithm [28–33]. Yang and Liu [28] proposed the physical incremental model by modeling the relationship between nodes of a network controlled by attraction force and repulsive force proposed in Newtonian mechanics. Other incremental dynamic community detection methods usually utilize the key feature of the network; these algorithms firstly obtain the community topology of network at the initial snap shot commonly using the static community detection method, then with time flying by, the variations of network during the man-set time range is recalculated and its community topology is identified [29–33]. However, these methods are usually designed for homogeneous networks.

2.2. Dynamic Community Detection of Heterogeneous Social Network. Lately, community detection in a heterogeneous network has become a hot research spot. Zhao et al. [34] proposed a uniform framework for detecting and tracking community evolution. They firstly model the entities and their relationships with the same timestamp into a heterogeneous network. They secondly extracted the snapshot-based feature and delta-based feature by utilizing the autoregression method to finally obtain the community topology and its evolution law. Sun et al. [35] introduced the community evolution in multimode networks and proposed a framework that partitioned the multimode network into a set of bipartie networks. Sun et al. used net clusters [14] to describe the community and proposed the Evo-NetCluster to detect the community automatically. Wu et al. [36] proposed a tensor decomposition framework to detect community in the general time-evolving heterogeneous network. Nevertheless, these methods either need to know the topology schema like star or bipartite or need to satisfy the requirements of tensor factorization, which are intractable/hard to use in real applications. Tang et al. [37] proposed a principal modularity maximization method, in which they first analyzed the modularity of different relational dimensions, then according to its eigenvalue and eigenvector of each relational dimension, the principal structural feature was extracted; thirdly, they correlated every network principal structural features to acquire the shared community topology of the whole network which would make the whole network modularity optimized.

With the rapid development of web 2.0 and mobile networks, event detection on heterogeneous data has drawn more attention in recent years. Yang et al. [38] proposed a

unified model to dynamically learn how to represent the data with different features of a heterogeneous social network. Liu et al. [39] treated the breaking news as a heterogeneous social data stream and developed how to extract events from the dynamic data stream. Liu et al. [40] extended the heterogeneous data stream into a multilingual scenario; Cao et al. [41] developed a knowledge-preserving and incremental social event detection framework using the GNNs and they applied it on the heterogeneous social network.

In summation, the TDT methods given above face three challenges as follows. Firstly, though most previous topic detection methods have good results in static online social networks, they rarely relate to the research of new emerging topic detection under the time-evolving and dynamic social network situations. Secondly, topic detection researchers mainly focus on finding new methods to detect prominent or distinct topics. They pay little attention to reveal the topic evolution process with time flying in the meantime of the topic detection process. Thirdly, the resolution limit problem has still not been well solved in the existing modularity community detection methods.

In this paper, we propose an original emerging topic detection and topology evolution identification framework to firstly detect the newly emerging online topic and secondly to uncover its evolution topology on the global heterogeneous online social networks.

3. Problem Formulation and Method

3.1. KeyGraph Network Modeling. Before introducing our proposed dynamic community detection method, we firstly build up the KeyGraph network for short texts crawled from heterogeneous social networks platforms in two steps. Firstly, every short text is modeled as a node/vertice of the key graph network. Connections between any two short texts are modeled as the edge state between them. Secondly, we acquire the keywords set corresponding to each short text by using the word segmentation technology. Thus the original short-text network can be abbreviated as a complex network based on the closeness of keywords.

In this paper, we denote and name it as KeyGraph $G = \{V_i, E_{ij}\}$ in the following way, where i, j represent the *ith* and *jth* short texts crawled from heterogeneous social networks and marked with a number, C_i is the keyword set of the *ith* short text using word segmentation technology, N_{ij} is the count number of common keywords belonging to keyword sets of both C_i and C_j; V_i is the *ith* node of the network, E_{ij} represents the edge between the *ith* and *jth* short texts which is closely related to the common keywords number N_{ij}. The relationship of E_{ij} with N_{ij} is shown in the following formula:

$$\begin{aligned} E_{ij} &= 1, \quad \text{if } N_{ij} > 0, \\ E_{ij} &= 0, \quad \text{if } N_{ij} = 0, w_{ij} = N_{ij}. \end{aligned} \tag{1}$$

For illustration purposes, 406 short texts containing both news titles and microblogs are crawled in which the number of people participating overcomes 1000 in October 1st, the year 2019. Its KeyGraph in the stochastic and Fruchterman Reingold distributions is shown in Figure 1. Figure 1(a) is the random distribution of the KeyGraph, and Figure 1(b) is the Fruchterman Reingold distribution of the KeyGraph. It shows a clear community structure.

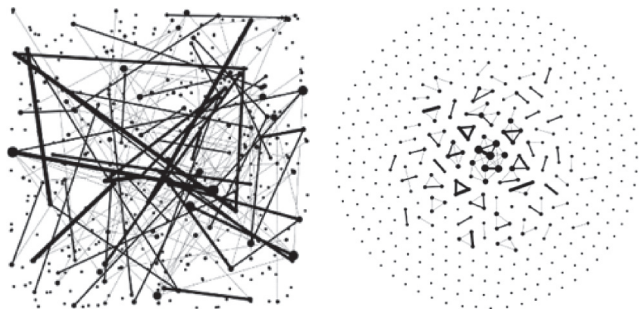
3.2. Dynamic Community Detection and Topic Detection on the Time-Evolving KeyGraph Network. Different from the static network, it should be noticed that the network formed by short texts crawled from heterogeneous online social network actually evolves with time flying by in this paper, and so do the relationships of the network. Thus the Keygraph network modeled in Section 3.1 is actually the time-evolving and scalable network. In this paper, we denote it as the time-evolving and scalable network $G_t = \{V_t, E_t\}$, the scale of which increases in sizes of either node V_t or edges E_t or even both of them with time flying by.

In this paper, we propose a dynamic community detection method which not only can effectively alleviate the resolution limit problem, but also can discover the community structure of the time-evolving and scalable network G_t. Its main idea is that at given time epoch t, the community structure of network at time epoch $t - \Delta t$, i.e., $G_{t-\Delta t}$ is assumed as clearly detected using static Louvain algorithm and already known, the part of network changing/variation during the time interval $(t, t + \Delta t]$ rather than the entire network, i.e., $G_{t+\Delta t}$ at time $t + \Delta t$ is our main focus. By calculating the closeness of local changed subnetwork during the time interval $(t, t + \Delta t]$ with historical communities of network G_t, a local bipartite graph is obtained. The local bipartite graph is composed of two groups of nodes. One is the group of nodes having loose closeness with communities of G_t and is denoted as $Bi\_G_{\text{loose},t,t+\Delta t}$, and the other is the group of nodes having close connections with communities of G_t denoted as $Bi\_G_{\text{close},t,t+\Delta t}$. By applying the static community detection, i.e., Louvain algorithm on subnetwork is composed of both the $Bi\_G_{\text{close},t,t+\Delta t}$ and historical communities of G_t; also by applying the Louvain algorithm on $Bi\_G_{\text{loose},t,t+\Delta t}$, the whole new emerging community loosely connected with historical communities of G_t will be detected. The community structure of time-evolving network at time epoch $t + \Delta t$, i.e., $G_{t+\Delta t} = \{V_{t+\Delta t}, E_{t+\Delta t}\}$ is discovered by combining community detection results on these bipartite graphs during the time range $(t, t + \Delta t]$. By simply utilizing the local changed network property, the complexity and running time only depend on the local changed part of network rather than the whole network at $t + \Delta t$, i.e., $G_{t+\Delta t} = \{V_{t+\Delta t}, E_{t+\Delta t}\}$, which enables its applications in the large-scale network. The flow chart of our proposed dynamic community detection method is presented in Figure 2.

Before explicitly unfolding the specific dynamic community detection method, we present some related and important definitions closely related to our algorithm in advance as follows.

3.2.1. Related Definitions

Definition 1. Closeness degree of node i with the network at time epoch t, i.e., G_t is defined as $k_{i,t}$ and is calculated using the following formula:



(a) (b)

Figure 1: Illustration of KeyGraph. (a) Random Distribution. (b) The Fruchterman Reingold distribution.

$$k_{i.t} = \sum_{V_{j_t} \in G_t} A_{ij_t}, \tag{2}$$

where A_{ij_t} is the adjacent matrix of the network $G_{t+\Delta t}$, j_t means node j belonging to network G_t, and node i belonging to network $G_{t+\Delta t}$.

Definition 2. Closeness of node i belonging to the local changed network of time range $(t, t + \Delta t]$ with networks G_t and $G_{t+\Delta t}$ is defined as r and is calculated using the following formula:

$$r = k_{i.t} - k_{i,t+\Delta t}, \tag{3}$$

if $r > 0$, it is believed that the node i during time range $(t, t + \Delta t]$ have a close relationship with the historical communities of network G_t compared with network $G_{t+\Delta t}$, if $r \leq 0$, then it is believed that node i during time range $(t, t + \Delta t]$ have a loose relationship with the historical communities of G_t compared with network $G_{t+\Delta t}$.

Definition 3. The modularity model proposed by Newman and Girvan is presented in the following formula:

$$Q = \frac{1}{2m} \sum_{i,j \in \Omega} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta_{ij}, \tag{4}$$

where A_{ij} is the adjacent matrix, δ_{ij} is the Kronecker function, $\delta_{ij} = 1$ when both nodes i and j are in the same community; otherwise, $\delta_{ij} = 0$. Ω is the set of total network nodes, k_i, k_j is the degree of nodes i and j within the whole network, m is the total weights of all edges of the whole network.

By rewriting the Kronecker function δ_{ij}, the modularity function Q can be rewritten in the following formula:

$$\begin{aligned} Q &= \frac{1}{2m} \left(\sum_{i,j \in \Omega} A_{ij} - \frac{\sum_{i \in \Omega} k_i \sum_{j \in \Omega} k_j}{2m} \right) \\ &= \frac{1}{2m} \sum C \left(\sum \text{in} - \frac{(\sum \text{tot})^2}{2m} \right), \end{aligned} \tag{5}$$

where C represents any community of network, $\sum$ in represents the total edge weights within community C, $\sum$ tot represents the summation of total edge weights connected with community C.

In our paper, a modularity gain index is defined as the modularity difference between the modularity value before and after reassigning node i into the community where node j belonging to the modularity gain is calculated using formulae (4) and (6).

$$\begin{aligned} \Delta Q_{i.j} &= \left[\frac{\sum \text{in} + k_{i.\text{in}}}{2m} - \left(\frac{\sum \text{total} + k_i}{2m} \right)^2 \right] \\ &\quad - \left[\frac{\sum \text{in}}{2m} - \left(\frac{\sum \text{total}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] = \frac{k_{i.\text{in}}}{2m} - \frac{k_i \sum \text{total}}{2m^2}. \end{aligned} \tag{6}$$

3.2.2. Dynamic Community Detection Method. After presenting the essential and necessary definitions and formulations, we give the specific framework of our dynamic community detection method for the time-evolving network as follows.

Firstly, for time-evolving network G_t formed by short texts and its relationships crawled from heterogeneous social network platforms before time epoch t, the static community detection algorithm here referred to as Louvain algorithm is utilized to obtain the community structure of G_t. Secondly, by bisecting the local varied network during the time range $[t, t + \Delta t]$ into two groups as a bipartite graph; one is the subnetwork denoted as $Bi\_G_{\text{close},t,t+\Delta t}$ composed of the new emerging nodes that have close relationship/connections with the historical communities of G_t; the other is the subnetwork denoted as $Bi\_G_{\text{loose},t,t+\Delta t}$ composed of the new emerging nodes that have loose relationship with the historical communities of network G_t, and we propose formula (3) to quantify the closeness of the local varied nodes within time range $[t, t + \Delta t]$ with historical communities of network G_t. Finally, by applying the static community detection method into the local varied networks $Bi\_G_{\text{loose},t,t+\Delta t}$ and $Bi\_G_{\text{close},t,t+\Delta t}$, we can identify the community subordinate attributes of these new emerging nodes within time range $[t, t + \Delta t]$, i.e., which nodes should belong to the historical communities of network G_t as the new emerging nodes within historical communities during time range $[t, t + \Delta t]$,

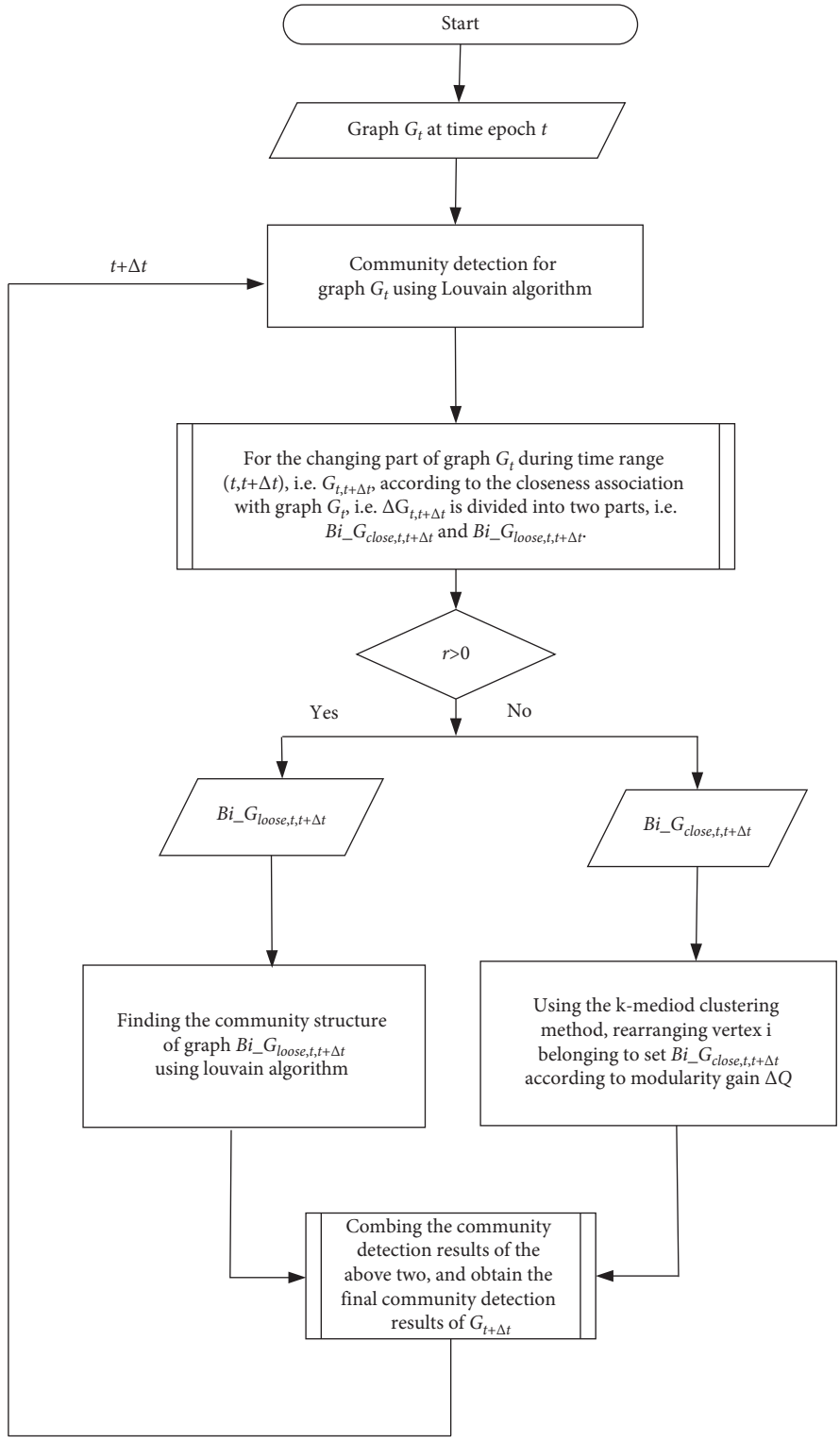


Figure 2: Flow chart of the community detection method based on the Louvain algorithm.

and which nodes should be assumed as the whole new emerging communities during time range $[t, t+\Delta t]$, respectively.

As mentioned above, there is a static community detection algorithm embedded and applied when we identify the community structure of bipartite networks $Bi\_G_{\text{close},t,t+\Delta t}$ and $Bi\_G_{\text{loose},t,t+\Delta t}$; here the Louvain algorithm is chosen as the embeded static community detection method. It is well known that a high value of modularity indicates a good community partition of the target network. Maximizing this criterion by utilizing all kinds of optimization algorithms has always been a popular research focus during the past

decades. However, it is intractable to find the exact global optimal modularity. Thus, many approximation optimization algorithms have been proposed. Among these algorithms, the greedy concept introduced by Blonde et al. [18] and called as Louvain algorithm has been proved to be among the most efficient algorithm with excellent performance, especially in large-scale networks. Louvain algorithm is actually the hierarchical cluster community detection method and mainly consists of two steps. In the first step, modularity is optimized locally in the neighborhood of each node; in the second step, it aggregates the nodes in the same community into supernodes and thus forms a new coarse-grained aggregated network. These two procedures are iteratively performed until the global value of network modularity stops to increase by any movement of nodes in the network, specific Louvain algorithm is presented as follows.

Step 1: treating each node of the targeted network as every single community

Step 2: for node i and its neighbor node j, we calculate the modularity gain $\Delta Q_{i,j}$ and its maximum value $\max \Delta Q_{i.j}$, if $\max \Delta Q_{i.j} > 0$, then we deem that node i and node j should belong to the same community

Step 3: repeating step 2 for all the node i and its neighbor node j until there is no community change for all the nodes of the network

Step 4: compressing the network with the community as an aggregated node, the degree of the aggregated node is the original degree of the corresponding community

Step 5: repeating step 1 to step 4 for the compressed network until the modularity gain $\Delta Q_{i.j}$ of the whole network does not increase, the algorithm stops

The modularity gain $\Delta Q_{i.j}$ in step 2 after node i joining into the communities of its neighbor node j is calculated using formula (6).

$$\Delta Q_{i.j} = \left[\frac{\sum \text{in} + k_{i.\text{in}}}{2m} - \left(\frac{\sum \text{total} + k_i}{2m}\right)^2\right] - \left[\frac{\sum \text{in}}{2m} - \left(\frac{\sum \text{total}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right] = \frac{k_{i.\text{in}}}{2m} - \frac{k_i \sum \text{total}}{2m^2}. \tag{7}$$

Thus, we have fully presented the dynamic community detection algorithm with the Louvain method embedded for the time-evolving network $G_t = \{V_t, E_t\}$.

Its main algorithm is presented as follows:

Step 1: Utilizing Louvain algorithm to obtain the community structure of network G_t

Step 2: bisecting the local varied network during time range $[t, t + \Delta t]$ into two groups as a bipartite graph using formula (3), denoted as $Bi\_G_{\text{close},t,t+\Delta t}$ and $Bi\_G_{\text{loose},t,t+\Delta t}$;

Step 3: applying the Louvain method for networks $Bi\_G_{\text{close},t,t+\Delta t} + G_t$ and $Bi\_G_{\text{loose},t,t+\Delta t}$, and combining the community detection results as the final community structure of network $G_{t+\Delta t}$.

3.2.3. Topic Detection Method. A more detailed process of our new emerging topic detection method is presented as follows. We first construct the named KeyGraph network using keywords of short texts crawled from heterogeneous social network platforms according to modeling rule shown in Section 3.1; secondly, by employing the time-evolving property of the constructed KeyGraph, we utilize the proposed dynamic community detection method to identify the community structure of time-evolving KeyGraph network; thirdly, for each detected community of KeyGraph, we calculate the total number of people participating in (i.e., reviewing, thumbing-up, retweeting) each detected community, which actually reflects the keywords of the original short texts belonging to the detected community; finally, we rank and select the top-N detected communities according to the sequence of the statistic value. According to the selected top-N communities, the highest frequently mentioned keywords of the detected community are chosen as the keywords of the newly detected emerging topic.

Until now, our topic detection method based on the dynamic community detection method has been fully uncovered, and its specific algorithm is presented as follows:

Step 1: the KeyGraph model is used to map the original short-text network into the KeyGraph network, thus the data set of the short texts before time epoch t is changing into the KeyGraph network G_t

Step 2: identifying the community structure of network G_t using Louvain algorithm

Step 3: adding the new emerging short-text data during time range $[t, t + \Delta t]$ into the network G_t and forming the new network at time epoch $t + \Delta t$, i.e., $G_{t+\Delta t}$

Step 4: identifying the community structure of network $G_{t+\Delta t}$ using the dynamic community detection method

Step 5: repeating step 3 and step 4 until the community structure of the short-text formed network $G_{t+\Delta t}$ has been fully identified

Step 6: calculate the total number of people participating in each detected community belonging to network $G_{t+\Delta t}$

Step 7: selecting the top-N communities which have the most statistical total number of people participating

Step 8: for the selected top-N communities, calculating the frequency of keywords subordinating to each community, chose the top-n keywords with the highest frequency as the keywords of the corresponding topic

3.3. Method Alleviating Resolution Limit Problem. Another advantage needs to be pointed out is that our proposed dynamic community detection method can effectively alleviate the resolution limit problem commonly encountered in modularity-based community detection methods of complex network. For illustration purposes,

Figure 3 is presented to show how our dynamic community detection method can alleviate the resolution limit problem by adaptively presetting the discrete time step $\Delta t_{0.i}, i = 1, 2, \ldots$ in the proposed dynamic community detection method.

As presented in Figure 3, at the initial time epoch t_0, we set the first discrete time step $\Delta t_{0.1}$, during the time range $[t_0, t_0 + \Delta t_{0.1}]$, the local changed networks are composed of nodes marked in orange circle and are denoted as community 1, light blue circles are denoted as community 2, and cyan circles are denoted as community 3; at time epoch $t_0 + \Delta t_{0.1}$, we adaptively set the second discrete time step as $\Delta t_{0.2}$, during time range $[t_0 + \Delta t_{0.1}, t_0 + \sum_{i=1}^{2} \Delta t_{0.i}]$, the corresponding local changed networks are composed of nodes marked as cyan circle, pink circle, and light blue circle, which are denoted as community 3, community 4, and community 2, respectively, where there are new emerging vertices in community 2 and community 3 compared with those communities at the previous time epoch $t_0 + \Delta t_{0.1}$, and community 4 is the whole new emerging community during time range $[t_0 + \Delta t_{0.1}, t_0 + \sum_{i=1}^{2} \Delta t_{0.i}]$; also at time epoch $t_0 + \sum_{i=1}^{2} \Delta t_{0.i}$, we set the time step $\Delta t_{0.3}$, during time range $[t_0 + \sum_{i=1}^{2} \Delta t_{0.i}, t_0 + \sum_{i=1}^{3} \Delta t_{0.i}]$, the local changed/varied networks are composed of nodes marked as pink circle, dark blue circle, where the scale of community 4 enlarges with new emerging vertices and in the mean time the whole new emerging community 5 is detected. Following this schema, this process continues in the nested and recursive way with time flying by. Thus, for the time-evolving and scalable network, by selecting the proper time step $\Delta t_{0.i}$, the resolution limit problem can be effectively alleviated.

3.4. Topic Evolution Law Identification. In this section, we focus on how to discover the community topology evolution of the detected topic in Section 3.2.2. It should be noticed that the constructed KeyGraph network is actually a directed graph under the topic propagation situation instead of simply topic detection scenario because the topic propagation commonly reflects information propagation direction during the topic diffuses process, while our topic detection method proposed in Section 3.2.2 is illustrated using the time-evolving, undirected KeyGraph network. Thus, in this section, we should first expand the modularity formula (4) for undirected networks into the directed network as shown in the formula.

$$Q = \frac{1}{n} \sum \left(a_{ij} - \frac{k_i^{\text{in}} k_j^{\text{out}}}{n} \right) \delta_{ij}, \tag{8}$$

where n is the total number of edges, a_{ij} is the i, j elements values of adjacent matrix of the directed network, k_i^{in} is the in-degree of node i, k_j^{out} is the out-degree of node j. δ_{ij} is the Kronecker function as defined before.

Also, we expand the modularity gain ΔQ for undirected graph into the directed graph and present it in the following formula:

$$\Delta Q = \frac{k_{i.\text{in}}}{n} - \frac{k_i \sum \text{tot}}{n^2}, \tag{9}$$

where k_i is the degree of node i, and $k_i = k_i^{\text{in}} + k_i^{\text{out}}$, $k_{i.\text{in}}$ represents a new number of edges connecting the local changed nodes with historical communities, $\sum \text{tot}$ represents the total number of edges with community C.

Hence, by substituting formulae (8) and (9) into our dynamic community detection method in Section 3.2.2, we can propose the algorithm for topic evolution topology identification purposes.

Step 1: modeling the users at time epoch t participating in the topic identified Section 3.2 as the directed topic propagation network G_t;

Step 2: identifying the community structure of the directed network G_t by utilizing our proposed dynamic community detection method;

Step 3: adding the local changed users participating into the directed topic propagation network G_t during time range $[t, t + \Delta t]$ and forming the scalable network $G_{t+\Delta t}$;

Step 4: separating the local changed users during time range $[t, t + \Delta t]$ into a bipartite graph, the one having a loose relationship with network G_t, denoted as $Bi\_G_{\text{loose},t,t+\Delta t}$, and the other having close relationship with network G_t, denoted as $Bi\_G_{\text{close},t,t+\Delta t}$;

Step 5: for subnetwork composed of G_t and $Bi\_G_{\text{close},t,t+\Delta t}$, employing the proposed dynamic community detection algorithm, the incremental information of the historical communities of G_t during time range $[t, t + \Delta t]$ is found.

Step 6: for subnetwork composed of $Bi\_G_{\text{loose},t,t+\Delta t}$, the whole new emerging community is identified by using the proposed dynamic community detection algorithm;

Step 7: merging the community detection results of steps 5 and 6, the community detection results of the topic propagation network at time epoch $t + \Delta t$ is identified;

4. Experiments and Results

To validate the effectiveness of our proposed topic detection and evolution law identification method, we use both the artificial complex networks composed of computer-generated data and the real network constructed by original data crawled from heterogeneous and popular social media platforms. By comparison of the community detection results under the static Louvain algorithm and the proposed dynamic community detection method with the Louvain algorithm embedded, our proposed dynamic community detection method yields better results which validate its effectiveness and feasibility.

4.1. Experiments on Artificial Computer-Generated Network. In the artificial complex network composed of computer-generated data, we generate the artificial complex networks by choosing the nodes' connection probability p within the same community, while the nodes connection probability between the communities is set as 1-p, and values p and 1-p satisfying $p > 1 - p$, which means that the closeness of nodes

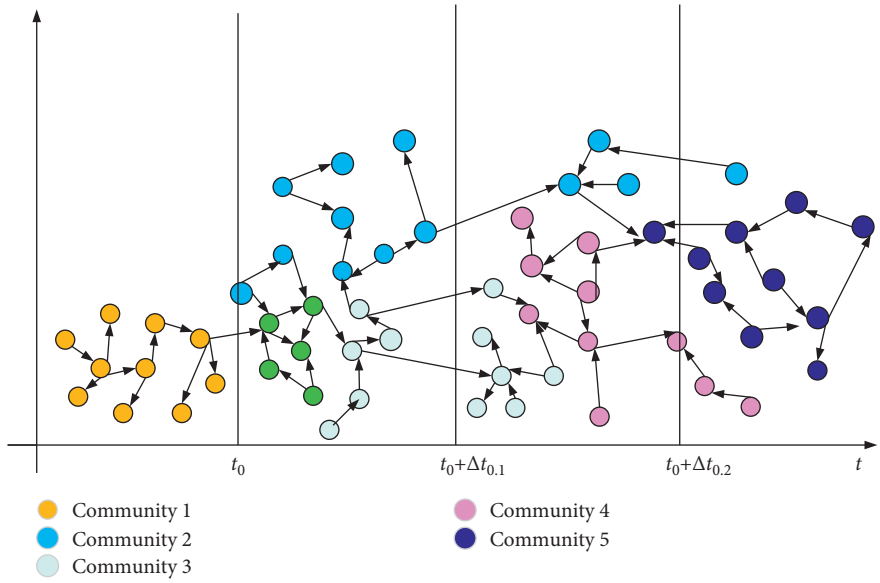


Figure 3: Illustration of dynamic community detection method to alleviate resolution limit problem.

within the community is larger than nodes between communities. Here, the artificial computer-generated network, which contains 68 nodes in total, with p set as 0.78. Community detection results using the static Louvain algorithm are presented in Figure 4(a).

While considering the time-evolving property in our practical situations, properties of nodes and edges of the artificial computer-generated network remains unchanged. While we randomly chose part of the nodes with time epoch t marked, those randomly selected nodes and edges at time epoch t compose the network G_t, and the remaining nodes are treated as the new emerging nodes of the network during time range $[t, t + \Delta t]$, the community detection results using our dynamic community detection methods are shown in Figure 4(b).

Figure 4 shows community detection results of the artificial computer-generated network with 68 nodes in total and edges randomly connected with parameter $p = 0.78$ described (a) using the static Louvain community detection algorithm and (b) using the proposed dynamic detection method with Louvain community detection algorithm, respectively. Nodes with the same color represent that they belong to the same community using corresponding community detection algorithms. The blue lines paralleling to the x-axes are the separating lines of different communities. The blue lines paralleling to the y-axes are the separating lines of time; as can be seen from Figure 4(b), the three blue lines paralleling to the x-axes separate the three different communities.

In order to verify the efficiency of the proposed dynamic community detection method with the Louvain algorithm embedded, we select five groups of artificial computer-generated data and use both the static Louvain algorithm and the dynamic community discovery method with the Louvain algorithm embedded to detect the communities structure of the computer-generated artificial network. Under the same operating environment, the time efficiencies of these two algorithms are compared and are shown in Figure 5.

It can be found from Figure 5 that the running time of the static Louvain algorithm is basically the same as those of the dynamic community discovery method with Louvain algorithm embedded when the network scale is relatively small (the number of vertices and edges of the network is relatively small), but with the fast growth of the network scale, the running time of our dynamic community detection method with Louvain algorithm embedded is far less than that of simple static Louvain community detection algorithm.

4.2. Experiments of Topic Detection on Real Short-Text Data Crawled From Real Social Networks Platforms. In this section, we randomly choose 86 short texts from October 1, 2019, to October 3, 2019, from heterogeneous online social networks for manual annotation. Later, the results of manual annotation are used to verify results validation of our proposed dynamic community detection method. Out of the 86 short texts, 46 short texts are crawled from Sina Weibo social platform, 17 short texts are crwaled from Sina News site social platform, 15 short texts are crawled from Sohu News site social platform, and 8 short texts are crawled from the Fenghuang News site social platform. In the results of manual annotations, the total 86 short texts are divided into 11 communities, out of which the largest community contains 28 pieces of short texts, and the smallest community contains 2 pieces of short texts.

According to the modeling rule of the KeyGraph network, the above-mentioned manual annotation short-text data are transformed into an undirected KeyGraph network, which has 86 nodes and 530 edges, among which 244 edges with a weight of 1 account for 46.04% of the total number of edges of the whole network. The weight range of the KeyGraph network is 0, 1, 2.

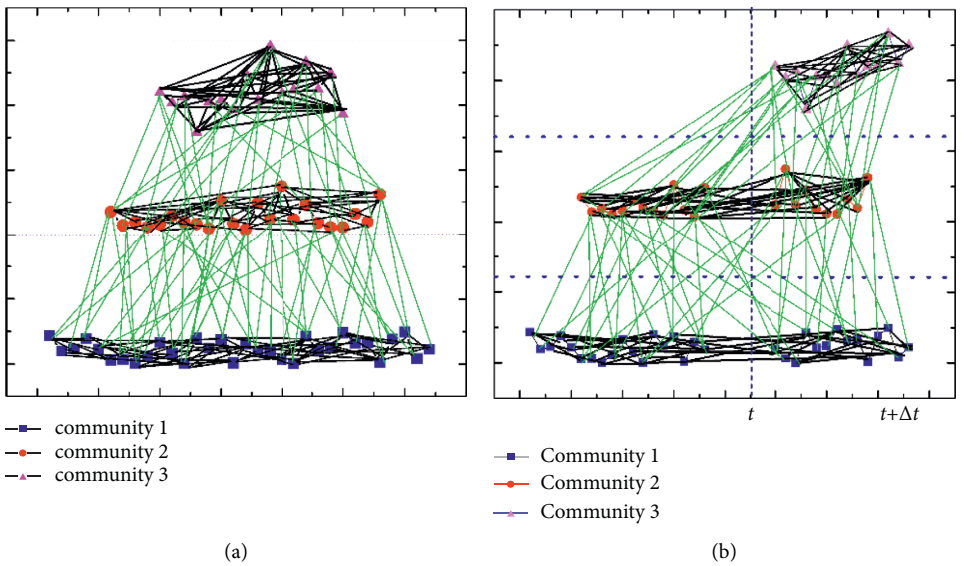


Figure 4: Community detection results of computer-generated complex network.

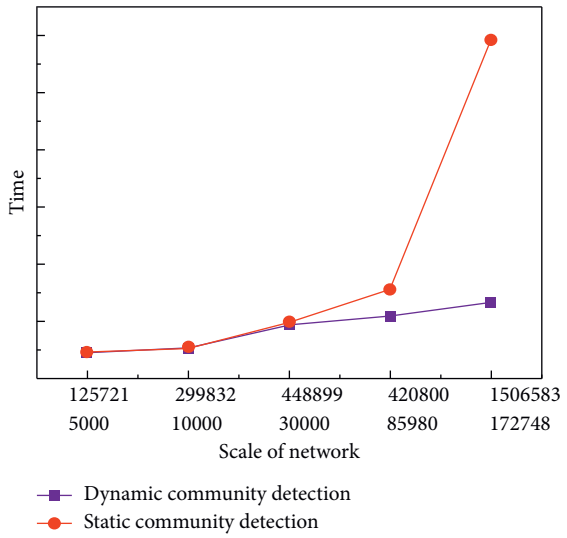


Figure 5: Performance comparison of the static Louvain algorithm and dynamic community detection method.

As we know that the weight of edge has a great effect on the final community detection results, and by applying the proposed dynamic community detection method with the Louvain algorithm embedded under different edge weights, the results are compared in Table 1.The detection ratio is computed using the following formula:

$$\text{Ratio} = \frac{n - s}{n}, \tag{10}$$

where n is the total number of manual annotation short texts, and s is the total number of misdetected communities.

From Table 1, it can be found that the community detection ratio of manual annotated short texts is higher, which means the community detection result is more accurate when the weight threshold value is set as 1 other than 0 and 2. Thus, in the topic identification experiments, we choose the edge weight threshold as 1.

Except for the pilot experiment to choose the optimal edge weight threshold, next, we will use the real short-text data crawled from the Sina Weibo, Sina News site, Sohu News site, and Fenghuang News site, the mainstream news publishing sites and microblog platforms which are popular. We media means citizen Journalism, the mainstream news publishing sites and microblog platforms which are popular citizen Journalism nowadays in China. A total of 262246 pieces of short-text data from October 1, 2019, to October 3, 2019, were extracted from the above listed heterogeneous online social networks as the real experimental data set.

According to the time stamp, the total number of crawled short texts on October 1 is 85980 pieces, the total number of crawled short texts on October 2 is 86768 pieces, and the total number of crawled short texts on October 3 is 89498 pieces. Chinese word segmentation and keywords extraction are performed on these crawled original experimental short-text

Table 1: Community detection ratio under different edge weights.

Weight	Community detected	Misdetected community	Detection ratio
0	10	3	0.965
1	12	1	0.988
2	24	22	0.744

data, and five keywords are selected to represent the original short texts for contents of news titles and microblogs.

Here, the time interval is set as 1 day (24 hours) and the edge weight threshold is set as 1 when we use the dynamic community detection method with the Louvain algorithm embedded. Firstly, we construct a network G_1 using the short-text data of October 1, and the network G_1 has 85980 nodes and 420800 edges based on the model definition of the KeyGraph network. Applying the proposed dynamic community detection method with Louvain algorithm embedded on KeyGraph network G_1. Secondly, adding the original crawled short-text data of October 2, then the network is denoted as KeyGraph network G_2, then the KeygGraph network G_2 has 172748 nodes and 1506583 edges, applying the proposed dynamic community detection method with Louvain algorithm embedded on KeyGraph network G_2; thirdly, adding the crawled original short-text data of October 3 into the KeyGraph network G_2, then the newly varied KeyGraph network is denoted as G_3, KeyGraph network G_3 has 262246 nodes and 3250235 edges. We apply the proposed dynamic community detection method with Louvain algorithm embeded on KeyGraph network G_3. The number of nodes corresponding to each community is shown in Figure 6, in which the abscissa represents the total detected community number and the ordinate represents the number of nodes belonging to each corresponding community.

The modularity of the KeyGraph network G_3 is 0.886, and 222133 communities are discovered by using the dynamic community detection method with the Louvain algorithm embedded; nodes contained in each corresponding community are shown in Figure 6. The abscissa represents the total detected number of communities, and the ordinate represents the number of nodes in each community.

From Figure 6, it can be found that almost all node sizes of the community are less than 250 and the proportion of community sizes less than 100 is larger than 99%. From the community discovery results, we can find the sparsity of the information distribution of the online social networks of different sources (222133 communities are found out of 262246 short texts), and the scale of the communities is generally small.

According to our proposed dynamic community detection method with Louvain algorithm embedded, we then calculate the total number of people participating in (replying, retweeting, thumbs-up) each short text (corresponding to each community); we rank it and choose the community in which the total number of people participants are larger than 100000; then we calculate the frequency of the keywords of each community, rank them, and obtain the top 5 keywords as the representative keywords of our detected topic, as shown in Table 2.

As shown in Table 2, the top 5 keywords of the 5 short-text communities are presented in the first column, the selected keywords are assumed as the keywords of topic detected as shown in the second column, and the total number of people participating in the short texts in ways either discussing, retweeting, or thumbs-up is shown in the third column.

It can be seen from Table 2, the top 5 keywords baring the highest appearing frequency in the short-text community are "Men's basketball, Asian championships, Iran, Chinese team, Asia;" they form the detected topic as "China Men's basketball Asian Championship;" the total number of people participating are the highest, about 391238 online users from heterogeneous social network platforms.

4.3. Validation of Topic Evolution Law. In this section, the detected topic "China men's basketball Asian Champion," which has the highest number of people participating, is chosen to validate our topic topology evolution identification purpose. 368 pieces of related news and microblogs are obtained, in which the participating people are 141318 in summation, out of which there are 1324 people on October 1, 8045 people on October 2, and 3872 people on October 3 in the year 2019.

After modeling the related data of the detected topic "China men's basketball Asian Champion" into the directed topic propagation network as the definition is shown in Section 3.4, there are 14318 nodes and 14962 edges in total. The community detection result of our dynamic community detection method on the directed topic propagation network is shown in Figure 7.

336 communities are detected and the modularity value is 0.514 by setting the time step as one day (24 hours). The dynamic properties of the topic propagation network is shown in Table 3. It can be found that until October 3, about 92.48% of users have already participated in the topic while about 56.19% users are added just on October 2 and about 258 on October 2 are new emerging communities.

From Figure 8, we can see that the scale of the node belonging to the community in each day corresponding to the detected topic "China Men's Basketball Asian Champion" reaches its peak at October 2 and gradually recedes to zero until October 6. The detected topic topology evolution is changing from October 1 as zero, reaches its peak value on October 2, and gradually recedes to zero from October 3 to October 6.

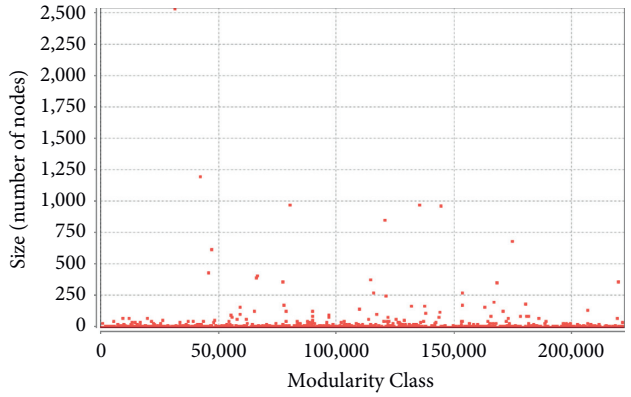


Figure 6: Size distribution of communities detected in the illustrated example.

Table 2: Topic detection results.

Top-5 keywords	Topic detected	People participating
Men's basketball, Asian championships, Iran, Chinese team, Asia,	China men's basketball Asian championship	391238
Liucheng county, Booming, Guangxi, package, express	The express package booms in Liucheng county, Guangxi	240673
Jewelry store, servant, Hongkong, customer, the worst	Hongkong jewelry store's servants provide the worst customer	195265
Commit adultery, condom, office, Hunan, Leiyang	Officer of Leiyang, Hunan commit adultery	167413
Innocent, Chinese, police, shoot, postgraduate	Japanese police shoot Chinese and is proved to be innocent	130957
National flag, broken, hung, county	A county in Jiangxi hang the broken national flag	123441

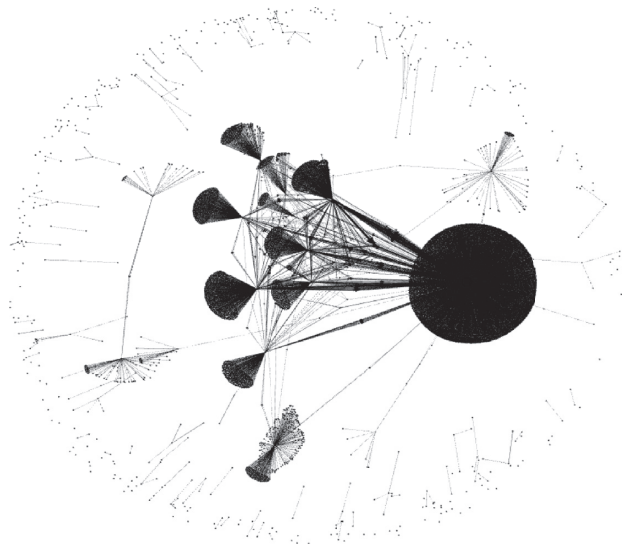


Figure 7: Community detection results of topic "China men's basketball Asian Champion."

Table 3: Properties of topic propagation network.

Day	Nodes	Edges	Value of modularity	Community detected
2019-10-01	1324	1451	0.558	12
2019-10-02	9369	9882	0.388	78
2019-10-03	13241	13881	0.457	336
2019-10-04	14318	14962	0.514	336

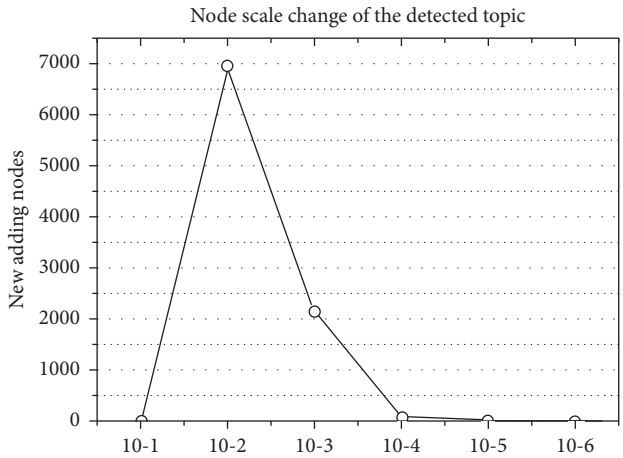


Figure 8: Variation of nodes in the detected topic "China men's Basketball Asian Champion."

5. Conclusions and Prospects

In this paper, we propose a topic detection and topology evolution identification framework based on the dynamic community detection method. Firstly, a unified time-evolving KeyGraph network is constructed based on the cooccurrence of keywords of short texts crawled from heterogeneous online social network platforms. Secondly, a dynamic community detection method for time-evolving network is proposed and the topic is detected by its utilization on the constructed KeyGraph network. Thirdly, for the detected topic in the previous step, a directional topic propagation network is built based on the short texts related to the detected topic, and the topic evolution topology is mainly reflected as the nodes scale of community is discovered.

Data Availability

The data used to support the finds of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] China Internet Network Information Center, *The 45th Development Statistic Report of Internet of China*, China Internet Network Information Center, Beijing, China, 2020, http://www.cnnic.cn/.

[2] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 1, pp. 7821–7826, 2002.

[3] E. M. Jin, M. Girvan, and M. E. J. Newman, "Structure of growing social networks," *Physical Review E-Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 64, Article ID 046132, 2001.

[4] A. Lancichinetti, M. Kivelä, J. Saramäki, and S. Fortunato, "Characterizing the community structure of complex networks," *PLoS One*, vol. 5, no. 8, Article ID e11976, 2010.

[5] J. Deng, K. Deng, Y. Li, and Y. Li, "Hot topic detection based on complex networks," in *Proceedings of the 2013 International Conference on Fuzzy Systems & Knowledge Discovery, IEEE*, Shenyang, China, July 2013.

[6] W. Lin and D. Guan-Zhong, "Forum hot topic detection based on community structure of complex networks," *Computer Engineering*, vol. 34, no. 11, pp. 214–217, 2008.

[7] Z. Liu and B. Hu, "Epidemic spreading in community networks," *Europhysics Letters*, vol. 72, no. 2, pp. 315–321, 2005.

[8] G. Ren and X. Wang, "Epidemic spreading in time-varying community networks," *Chaos An Interdiplinary Journal of Nonlinear Science*, vol. 24, no. 2, p. 068701, 2014.

[9] X. Wang, L. Tang, H. Liu, and L. Wang, "Learning with multi-resolution overlapping communities," *Knowledge and Information Systems*, vol. 36, no. 2, pp. 517–535, 2013.

[10] L. Tang, X. Wang, and H. Liu, "Community detection via heterogeneous interaction analysis," *Data Mining and Knowledge Discovery*, vol. 25, no. 1, pp. 1–33, 2012.

[11] M. Revelle, C. Domeniconi, M. Sweeney et al., "Finding community topics and membership in graphs," in *Joint European Conference on Machine Learning & Knowledge Discovery in Databases*Springer International Publishing, New York, NY, USA, 2015.

[12] Y. Hu, B. Yang, and C. Lv, "A local dynamic method for tracking communities and their evolution in dynamic networks," *Knowledge-Based Systems*, vol. 110, no. 1, pp. 176–190, 2016.

[13] R. Aktunc, I. H. Toroslu, M. Ozer, and H. Davulco, "A dynamic modularity based community detection algorithm for large-scale networks: DSLM," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Paris, France, August 2015.

[14] Y. Sun, Y. Yu, and J. Han, "Ranking-based clustering of heterogeneous information networks with star network schema," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, Paris, France, June 2009.

[15] M. D. Hoffman, D. M. Blei, and F. R. Bach, "Online learning for latent dirichlet allocation," in *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Vancouver, British Columbia, Canada, December 2010.

[16] Y. Akita, Y. Nemoto, and T. Kawahara, "PLSA-based topic detection in meetings for adaptation of lexicon and language

model," in *Proceedings of the Interspeech, Conference of the International Speech Communication Association*, Antwerp, Belgium, August 2007.

[17] H. Sayyadi and L. Raschid, "A graph analytical approach for topic detection," *ACM Transactions on Internet Technology*, vol. 13, no. 2, pp. 1–23, 2013.

[18] V. D. Blonde, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10, no. 1, p. 10008, 2008.

[19] R. Jin, C. Kou, and R. Liu, "Improving community detection in time-evolving networks through clustering fusion," *Cybernetics and Information Technologies*, vol. 15, no. 2, pp. 63–74, 2015.

[20] T. Murata and S. Moriyasu, "Blog community discovery and evolution based on mutual awareness expansion," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, Fremont, CA, USA, November 2007.

[21] A. Cuzzocrea, F. Folino, and C. Pizzuti, "Dynamicnet: an effective and efficient algorithm for supporting community evolution detection in time-evolving information networks," in *Proceedings of the 17th International Database Engineering and Applications Symposium (IDEAS '13)*, Barcelona, Spain, October 2013.

[22] Y. Lin, C. Yun, S. Zhu, H. Sundaram, and B. L. Tseng, "Analyzing communities and their evolutions in dynamic social networks," *ACM Transactions on Knowledge Discovery from Data*, vol. 3, no. 2, 2008.

[23] F. Folino and C. Pizzuti, "An evolutionary multiobjective approach for community discovery in dynamic networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1838–1852, 2014.

[24] L. Tang, H. Liu, J. Zhang et al., "Community evolution in dynamic multimode networks," in *Proceedings of the 14th Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*, Las Vegas, NV, USA, August 2008.

[25] M. Toyoda and M. Kitsuregawa, "Extracting evolution of web communities from a series of web archives," in *Proceedings of the 14th ACM Conference on Hypertext and Hypermedia*, pp. 28–37, Nottingham, UK, August 2003.

[26] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 3, pp. 814–818, 2005.

[27] D. Chakrabarti, R. Kumar, and A. Tomkins, "Evolutionary clustering," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 554–560, Philadelphia, PA, USA, August 2006.

[28] B. Yang and D.-Y. Liu, "Force-based incremental algorithm for mining community structure in dynamic network," *Journal of Computer Science and Technology*, vol. 21, no. 3, pp. 393–400, 2006.

[29] H. Ning, W. Xu, Y. Chi, Y. Gong, and T. S. Huang, "Incremental spectral clustering by efficiently updating the eigensystem," *Pattern Recognition*, vol. 43, no. 1, pp. 113–127, 2010.

[30] C. Dhanjal, R. Gaudel, and S. Clémençon, "Incremental spectral clustering with the normalized laplacian," in *Proceedings of the 3rd NIPS Workship on Discrete Optimization in Machine Learning*, pp. 1–6, Granada, Spain, December 2011.

[31] D. Duan, Y. Li, R. Li, and Z. Lu, "Incremental k-clique clustering in dynamic social networks," *Artificial Intelligence Review*, vol. 38, no. 2, pp. 129–147, 2011.

[32] T. N. Dinh, N. P. Nguyen, and M. T. Thai, "An adaptive approximation algorithm for community detection in dynamic scale-free networks," in *Proceedings of the IEEE International Conference on Computer Communications*, pp. 55–59, Turin, Italy, April 2013.

[33] T. Falkowski, A. Barth, and M. Spiliopoulou, "Studying community dynamics with an incremental graph mining algorithm," in *Proceedings of the 14th Americas Conference on Information Systems*, p. 29, Toronto, Cananda, August 2008.

[34] Q. Zhao, S. Bhowmick, X. Zheng, and Y. Kai, "Characterizing and predicting community members from evolutionary and heterogeneous networks," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pp. 309–318, Napa Valley, CA, USA, October 2008.

[35] Y. Sun, J. Tang, J. Han, M. Gupta, and B. Zhao, "Community evolution detection in dynamic heterogeneous information networks," in *Proceedings of the 8th Workshop on Mining and Learning with Graphs (MLG '10)*, Washington, D.C., USA, July 2010.

[36] J. Wu, Y. Lianfei, Z. Qun et al., "Multityped community discovery in time-evolving heterogeneous information networks based on tensor decomposition," *Complexity*, vol. 2018, Article ID 9653404, 16 pages, 2018.

[37] L. Tang, X. Wang, and H. Liu, "Uncovering groups via heterogeneous interaction analysis," in *Proceedings of the 2009 International Conference on Data Mining*, pp. 503–512, Miami, FL, USA, December 2009.

[38] C. Yang, Y. Xiao, Y. Zhang, Y. Sun, and J. Han, "Heterogeneous network representation learning: survey, benchmark, evaluation, and beyond," 2020, http://arxiv.org/abs/2004.00216.

[39] B. Liu, F. X. Han, D. Niu, L. Kong, K. Lai, and Y. Xu, "Story forest," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 3, pp. 1–28, 2020.

[40] Y. Liu, H. Peng, J. Li et al., "Event detection and evolution in multi-lingual social streams," *Revue Finance, Contrôle, Stratégie*, vol. 14, no. 5, pp. 1–15, 2020.

[41] Y. Cao, H. Peng, J. Wu et al., "Knowledge-preserving incremental social event detection via heterogeneous GNNs," in *Proceedings of the Web Conference 2021*, Ljubljana, Slovenia, April 2021.

Hindawi
Complexity
Volume 2021, Article ID 5526412, 13 pages
https://doi.org/10.1155/2021/5526412

Research Article

Anonymous Authentication and Key Agreement Scheme Combining the Group Key for Vehicular Ad Hoc Networks

Mei Sun,[1,2] **Yuyan Guo**,[2] **Dongbing Zhang**,[2] **and MingMing Jiang**[2]

[1]*School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China*
[2]*School of Computer Science and Technology, Huaibei Normal University, Huaibei 235000, China*

Correspondence should be addressed to Mei Sun; sunmei109@163.com

Received 24 January 2021; Revised 25 March 2021; Accepted 20 April 2021; Published 4 May 2021

Academic Editor: Jia Wu

Vehicular ad hoc network (VANET) is a multihop mobile wireless communication network that can realize many vehicle-related applications through multitop communication. In the open wireless communication environment, security and privacy protection are important contents of VANET research. The most basic method of VANET privacy protection is anonymous authentication. Even through, there are many existing schemes to provide anonymous authentication for VANETs. Many existing schemes suffer from high computational cost by using bilinear pairing operation or need the assistance of the trust authorities (TAs) during the authentication process or rely on an ideal tamper-proof device (TPD), which requires very strong security assumption. In this study, an anonymous authentication and key negotiation scheme by using private key and group key is proposed, which is based on pseudonym using the nonsingular elliptic curve. In this scheme, there is no third party trust center to participate in the authentication, there is no need to query the database, and there is no need of the local database to save the identity information of many vehicles, which reduce the storage space and the authentication time compared with other schemes. The proposed scheme only needs realistic TPDs. In the proposed scheme, TPDs do not need to preinstall the system key as many other schemes do; hence, the failure of a single TPD does not affect the security of the entire system. The security of the scheme is proved under the random oracle model. Compared with the related schemes using bilinear pairings, the computational cost and communication cost of the proposed scheme are reduced by 82% and 50%, respectively.

1. Introduction

With the development of network technology, there are many forms of network and new technologies [1, 2]. Vehicular ad hoc network (VANET) is a highly mobile self-organizing wireless communication network. By using VANET, vehicles in front can in a timely manner report the road condition information to the rear vehicles; this can improve the travel efficiency and reduce road congestion and traffic accidents. VANET plays a significant role in traffic optimization and safety [3]. Since VANET mainly adopts a wireless communication mode, messages are vulnerable to various attacks, such as counterfeiting, interception, tampering, tracking, and other attacks [4, 5]. These attacks seriously threaten the safety of vehicles and the privacy of users. Therefore, security authentication and privacy protection are important research directions of VANET. VANET generally has the following main components: road side unit (RSU), trust agency (TA), and on-board unit (OBU) [6]. OBU is installed in the vehicle and can realize the communication between the vehicle and RSU or other vehicles. The communication between OBU and RSU adopts dedicated short range communication (DSRC) [7]. The communication with vehicles requires authenticating one another and negotiating the communication key to prevent attacks such as tracking, privacy exposure, and message counterfeiting. Authentication and key agreement in VANETs are anonymous. Hence, even if an attacker intercepts the message, the specific source of the message cannot be determined. Additionally, the authority of

VANET can identify every message sent by vehicles, and this can prevent vehicles from sending false messages maliciously.

1.1. Related Works. In recent years, some authentication protocols based on public key infrastructure (PKI) [8–10] have been proposed. In these works, some anonymous authentication and key agreement schemes are proposed, in which a large number of certificates are assigned to vehicles. However, these schemes require vehicles to be equipped with many anonymous certificates in advance; this leads to many problems such as certificate storage and certificate management. Lu et al. [11] proposed a key agreement and authentication scheme for generating a short-term key and certificate between the vehicle and RSU. However, the communication efficiency of the scheme is low due to the frequent interaction between the vehicle and RSU for changing the authenticated group. Rajput et al. [12] proposed an anonymous authentication scheme with hierarchical privacy protection to solve the defects based on PKI. This protocol does not need to manage the certificate revocation list (CRL), and each vehicle uses two pseudonyms to complete anonymous authentication, but once the pseudonym expires, the vehicle needs to acquire the pseudonym from TA or RSU again; this increased the number of communications. Wang [13] proposed a local identity-based anonymous authentication protocol for VANET (LIAP). In this method, each vehicle and RSU are assigned a unique long-term certificate from the certification authority (CA) in the registration phase. The vehicle and RSU complete mutual authentication through certificates. After successful authentication, RSU distributes a local-master key to the vehicle. The vehicle randomly generates a pseudonym to communicate with the RSU through the local-master key. The use of the local-master key improves the communication efficiency and system security. But this scheme needs to manage CRL.

The storage and management of certificates restrict the development of authentication schemes based on PKI. To overcome the problems caused by authentication certificates, some identity-based public key cryptosystems are introduced into authentication of VANET [4, 14–19]. In 1984, Miller first proposed an identity cryptosystem [14]. In this cryptosystem, the user's public key is calculated by the user's identity, and the user's private key is generated by the authentication center through the system key according to the user's identity. In 2008, Zhang et al. [15] proposed an authentication protocol for VANET using the identity of the vehicle user, solving the certificate storage and management problem and supporting batch authentication. In 2011, Huang [16] proposed an anonymous batch authenticated and key agreement scheme based on identity authentication for VANET. Shim et al. [17] noted that the scheme [15] was vulnerable to replay attack and did not achieve the nonrepudiation of signature and proposed a vehicle-to-infrastructure (V2I) authentication scheme. However, the scheme is vulnerable to tampering attacks [18] and cannot satisfy its claimed chosen message attack resistance [17]. Wang et al. [20] mentioned that Huang et al. [16] could not resist a collusion attack, and therefore, they proposed an improved scheme. And, in [20], it is indicated that the scheme [18] cannot resist replay attacks and cannot track the real identity of the message sender. In 2016, Azees and Vijayakumar [21] proposed a novel key distribution scheme for secure group communication using Lagrange polynomials. The limitation of the scheme is that it only provides one-way authentication from vehicle to TA. Then, Vijayakumar et al. [22] proposed a privacy-preserving anonymous mutual and batch authentication scheme for vehicle-to-vehicle. This scheme implements the authentication of message source and message integrity and has the mechanism of tracking and revoking vehicles. In 2017, Azees et al. [23] proposed an anonymous authentication scheme to avoid malicious vehicles into the VANET based on bilinear pairing. Each user computes multiple temporary short time certificates to realize anonymous authentication in the scheme. The scheme has high computing performance and security. However, the dummy identity (DIU_{ui}) in each certificate is the same, and the scheme does not consider the unlinkability of different sessions. In 2018, Pournaghi et al. [24]. proposed an anonymous authentication and key agreement scheme combining TPD and RSU. The scheme saves the system master key in the TPD of RSU instead of the TPD of each vehicle, which improves the security and authentication efficiency of the system. In 2019, Ikram et al. [25] proposed a conditional privacy-preserving authentication scheme for V2I. This scheme uses general one-way hash functions instead of map-to-point hash functions to achieve high efficiency.

The identity-based authentication schemes for VANET address the problems presented by the schemes based on PKI. The existing schemes [21–25] are novel in design and have good security. However, the bilinear pairing operations of elliptic curve are used, and the computational efficiency of bilinear pairing operation is low. The works [26–28] based on pseudonym on elliptic curve, which do not use bilinear pairing operation and have achieved high computational efficiency. However, TA is required to participate in authentication, this increases communication times and communication burden. He et al. [29] proposed a privacy protection authentication scheme based on identity. This scheme also uses elliptic curve instead of bilinear pairing operations and achieves satisfactory performance in both computation and communication. However, the scheme is based on ideal TPD, and the master key is stored on the TPD of each vehicle. Islam et al. [30] proposed a conditional privacy-preserving authentication scheme based on hash function. And the scheme offers group-key generation, user leaving, user join, and password change facilities. The scheme does not need bilinear pairing mapping or elliptic curve operation and is lightweight in terms computation and communication. However, TA is required to participate in each authentication between the vehicle and RSU. Wu et al. [6] proposed an effective location-based conditional secret authentication scheme. The scheme does not require bilinear pairing operations or TPDs. However, when RSU is certified by vehicle, TA needs to query the database and return the

results. Cui et al. [31] proposed a scheme without relying on any special hardware such as TPD. The scheme is based on elliptic discrete logarithm and has high computational performance. The cuckoo filter and binary tree search method are used to achieve a higher success rate in batch authentication. However, TA is required to generate communication key for the vehicle and RSU. Zhong et al. [32] proposed an authentication and key agreement scheme based on hash function and registration list. And the scheme does not require the strong security assumptions of TPD. Xiong Li et al. [33] proposed a lightweight authentication scheme for VANETs with only hash functions and exclusive-OR operations. Compared with previous schemes, the computational cost of the schemes [32, 33] has been greatly improved. However, the schemes also need TA to participate in the authentication. In recent years, there are some authentication schemes using group key, which can reduce the authentication burden of TA. The works [34, 35] introduce group key management schemes based on Chinese remainder theorem, which reduces computation complexity of the key server. In 2019, Jing Zhang et al. [36] proposed a message authentication scheme based on the group key using Chinese remainder theorem. The TPD of the vehicle only save the real identity and the group key. So the proposed scheme only requires realistic TPDs and ensures higher security for the entire system. In 2020, Wei et al. [37] proposed tow privacy-preserving multimodal implicit authentication protocols for Internet of connected vehicles. The proposed protocols use the password and vehicle owner's behavior features as the authentication factors skillfully and do not reveal any information about vehicle owner's behavior. The protocols have advantages in computational cost and accuracy. However, the protocols do not consider the unlinkability of sessions. Vinoth et al. [38] proposed a multifactor authenticated key agreement scheme for industrial Internet of things (IoT). The scheme implements authentication and key agreement between the user and multiple sensing devices at the same time. The scheme only used hash function, bit-wise XOR operation, and symmetric cryptography. It has less communication cost and computational cost compared with other correlative schemes. However, the scheme does not consider internal attack.

1.2. Our Contributions. In this study, an anonymous authentication and key agreement scheme based on elliptic curve for VANET is proposed. Each vehicle is equipped with a TPD. The TPD saves the private key of the vehicle and the group key for multivehicle communication. The vehicle can authenticate with RSU anonymously by combining a private key with a group key. After successful authentication, the session key can be negotiated for both parties. The scheme can also implement message signature and anonymous verification. In this scheme, the TPD only saves the private key of the vehicle and the group key instead of the system key. The attack on the TPD will not affect other nodes in VANET. So, we only need realistic TPD instead of ideal TPD. There is no need for the third party to participate in the authentication and key agreement between vehicle and RSU compared with the works [6, 30–33], and there is no need to query the database in the scheme. In addition, the use of group key in this scheme can help RSU resist certain denial of service (DoS) attacks.

The main contributions of this study are summarized as follows.

(1) In order to optimize the computational cost and key management, we present an efficient anonymous authentication and key agreement scheme for RSUs and vehicles using the private key of the vehicle and the group key

(2) In order to reduce the communication time and storage space, we implement independent authentication and key agreement between vehicle and RSU, and RSU does not need to save vehicle information or query database.

(3) In this scheme, we also implement anonymous signature and verification of messages

(4) In this scheme, we use realistic TPDs instead of ideal TPDs, which is more suitable for VANET

1.3. Organization of This Article. The rest of the study is structured as follows: Section 2 describes the preliminaries of the proposed scheme, Section 3 gives the working of the proposed scheme, Sections 4 and 5 present a security analysis and a performance analysis, respectively. Our study is concluded in Section 6.

2. Preliminaries

In this section, we introduce the related background information of VANET and the proposed scheme.

2.1. Network Model. As shown in Figure 1, the network model of VANET mainly includes TA, RSU, OBU, TPD, and application server (AS). TA is a trusted service center. It is responsible for generating the private and public keys for RSU and vehicle and the group key for multivehicle communication. TA is an entity with the highest level of security protection and is completely trusted. RSU is the communication equipment installed on both sides of the road, with high security, thus providing access service for vehicles. The RSU communicates with the vehicle using DSRC protocol. Each vehicle is equipped with an OBU. The OBU of the vehicle realizes short distance communication with RSU and OBUs of other vehicles. TA allocates a TPD to each vehicle. TPD has high security, and other attackers cannot obtain sensitive information from the device [39]. AS is an application server and provides data service for TA. AS has high security and is credible.

2.2. Elliptic Curve. Suppose that F_P denotes a finite field of order p, where p is a large prime number. E denotes an elliptic curve over F_P. The curve E is defined as $y^2 = x^3 + ax + b \bmod p$, where $a, b \in F_p$. The group G is a

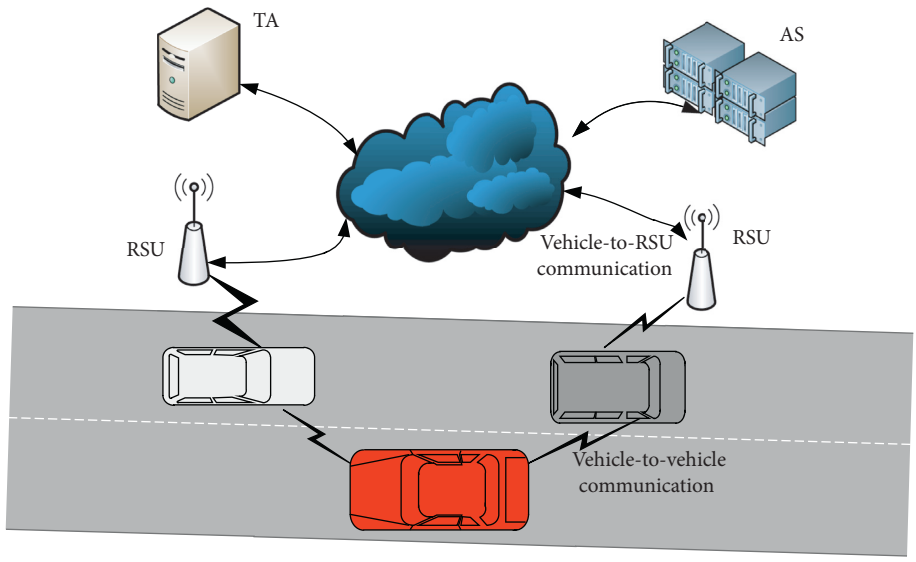


Figure 1: VANET network model.

cyclic additive group of order q on E, and P is the generator and O is the infinite point.

The group G has the following properties:

(1) Additive (±). For $P, Q \in G$, if $P \neq Q$, $R = P + Q$, then R is the intersection point of the straight line passing through P and Q with E; if $P = Q$, $R = P + Q$, then R is the tangent intersection point of P and Q with E; if $P = -Q$, then $P + Q = P - P = O$.

(2) Scalar multiplication (.). Let $m \in Z_q^*$, scalar point multiplication in G is defined as $m.P = P + P + \ldots + P$ (m times).

Two difficult problems are defined as follows:

Definition 1. Elliptic curve discrete logarithm problem (ECDLP). Let Q be a random point on G and calculate a solution x which satisfies $Q = xP$, where $x \in Z_q^*$.

Definition 2. Elliptic curve computational Diffie–Hellman problem (ECCDH). Assume a generator P of G, $aP, bP \in G$, where $a, b \in Z_q^*$ are unknown. The ECCDH problem is to compute $abP \in G$.

If ECDLP or ECCDH on a group G cannot be solved with nonnegligible probability ε in time t, then ECDLP or ECCDH is said to be a difficult problem on elliptic curve.

2.3. Security Requirements. The open multihop wireless network is vulnerable to various attacks. Therefore, the authentication and key agreement for VANET need to meet the following security requirements [29, 39]:

(1) Authentication and integrity. After receiving the message, VANET needs to determine whether the source of the message is reliable and whether the message has been tampered by others

(2) Privacy protection. When users are communicating, VANET should protect the confidential information such as user's identity, session record, location, and driving path. VANET provides privacy protection by imparting anonymity.

(3) Session key agreement. When the vehicle transmits data with RSU, the session key should be used to encrypt the data to protect the session privacy

(4) Traceability. To prevent malicious users from sending false messages by anonymity, the authentication scheme should trace the real identity of the sender when the message is in dispute

(5) Resistance to attacks. VANET is vulnerable to various attacks, such as replay attacks and forgery attacks. Authentication and key agreement of VANET needs to be able to resist all kinds of attacks to ensure the security and reliability of the scheme.

(6) Unlinkability. In order to protect privacy, attackers or other vehicles cannot link different sessions of the same vehicle via the public channel.

3. Proposed Authentication Scheme for VANET

Our scheme includes the following phases: initialization, RSU and vehicle registration, authentication and communication key agreement, message signing, signature verification, identity extraction, and updating the group key. The mutual authentication and the key agreement process between RSU and the vehicle is shown in Figure 2. The main notations used in the scheme are given in Table 1.

3.1. Initialization Phase. TA selects random numbers $s, x \in Z_q^*$, s is the private key of the system, x is the group key for multivehicle communication, and it can be used to compute the public key $P_{\text{pub}} = sP \in G$. Furthermore, $P_x = xP \in G$. TA selects five secure hash functions: $h_0: \{0,1\}^* \times G \longrightarrow Z_q^*$, $h_1: G \longrightarrow Z_q^*$, $h_2: \{0,1\}^* \times \{0,1\}^* \times G \times \{0,1\}^* \longrightarrow Z_q^*$, $h_3: \{0,1\}^* \times \{0,1\}^* \times G \times G \times \{0,1\}^*$, h_4

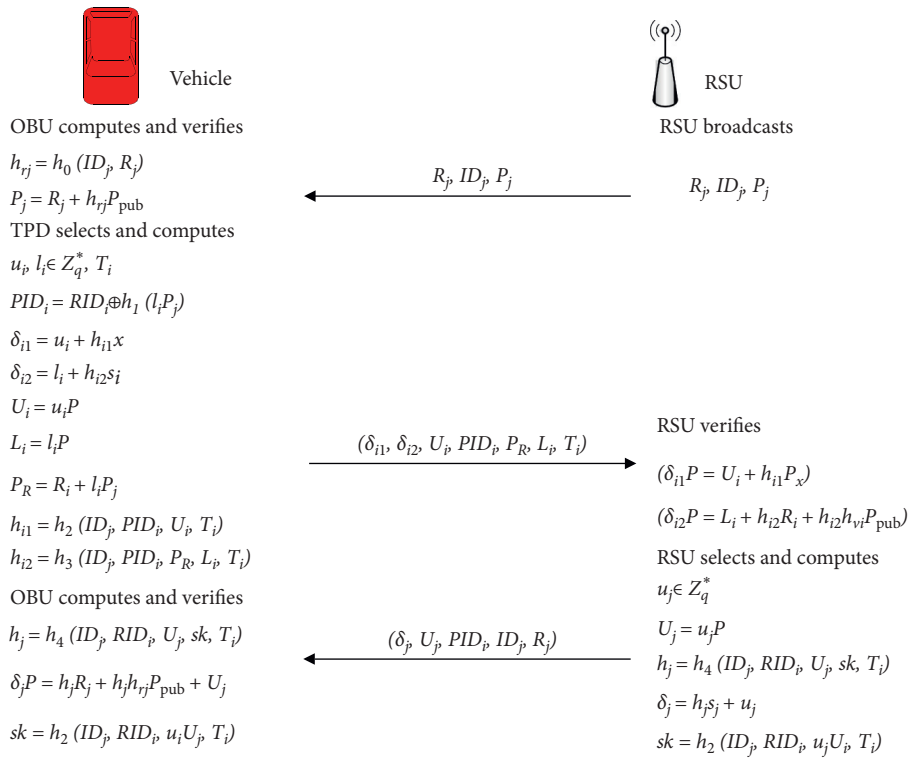


Figure 2: Mutual authentication and key agreement.

Table 1: Notations used.

Notation	Description
E	An elliptic curve
G	An additive group based on E
P	A generator of G
p, q	Large prime numbers
s, P_{pub}	Private key and public key pairs of the system
X, P_x	Group key and group public key pairs
h_0, h_1, h_2, h_3, h_4, h_5	Six secure hash functions
ID_j	Identity of the RSU
s_j, P_j	Private key and public key pairs of the RSU
s_i, P_i	Private key and public key pairs of the vehicle
RID_i	Real identities of the vehicle
PID_i	Pseudonym of the vehicle
T_i, T_m	Timestamp
sk	Session key between RSU and the vehicle
M_i	Traffic-related message

: $\{0,1\}^* \times \{0,1\}^* \times G \times \{0,1\}^* \times \{0,1\}^* \longrightarrow Z_q^*$, and h_5: $\{0,1\}^* \times \{0,1\}^* \times G \times G \times G \times \{0,1\}^* \longrightarrow Z_q^*$. TA also broadcasts the system parameters: Paras = $\{E, a, b, p, q, P, P_{pub}, P_x, h_0, h_1, h_2, h_3, h_4, h_5\}$.

3.2. RSU and Vehicle Registration Phase. Roadside unit RSU_j applies to TA for registration. After TA verifies the information of RSU_j successfully, it allocates the identity ID_j to RSU_j. Then, TA selects a random number r_j, computes $h_{rj} = h_0(ID_j, R_j)$ and $R_j = r_jP$. TA also generates the private key $s_j = r_j + h_{rj}s$ and then returns R_j, $s_j$to RSU_j. RSU_j computes $P_j = s_jP$ and verifies whether the following equation holds.

$$P_j = R_j + h_{rj}P_{pub},$$
$$\because P_j = s_jP = r_jP + h_{rj}sP = R_j + h_{rj}P_{pub}. \quad (1)$$

If (1) holds, RSU_j broadcasts R_j, ID_j, and P_j. Otherwise, the message is rejected. After RSU broadcasts the public key P_j, the vehicle can use P_j to compute the pseudonym of the vehicle. The detailed process is shown in Section 3.3.

During the registration process, the vehicle users go to TA directly. The vehicle users submit the required information such as identification, phone number, and license, etc., to TA. TA checks whether the vehicle user is qualified. If the vehicle user is qualified, TA allocates a TPD to the vehicle V_i and assigns a unique identity RID_i to the vehicle V_i. TA allows users to set a username and password for TPD. Then, TA chooses a random number r_i and computes $R_i = r_iP$, $h_{vi} = h_0(RID_i, R_i)$, $s_i = r_i + h_{vi}s$, and $P_i = s_iP$. TA saves s_i, x, RID_i, R_i, and P_i in the TPD of the vehicle V_i. At the same time, the vehicle information such as RID_i, R_i, and P_i is saved in AS.

3.3. Authentication and Communication Key Agreement Phase. RSU_j broadcasts R_j, ID_j and P_j; the OBU of the vehicle receives them and verifies whether (1) holds. If it holds, the OBU forwards them to the TPD of the vehicle. The TPD selects the random numbers $u_i, l_i \in Z_q^*$, and the timestamp T_i. The TPD computes the pseudonym$PID_i = RID_i \oplus h_1(l_iP_j)$ and generates the signatures $\delta_{i1} = u_i + h_{i1}x$

and $\delta_{i2} = l_i + h_{i2}s_i$, where $U_i = u_iP$, $h_{i1} = h_2(ID_j, \mathrm{PID}_i, U_i, T_i)$, $L_i = l_iP$, $P_R = R_i + l_iP_j$, and $h_{i2} = h_3(\mathrm{ID}_j, \mathrm{PID}_i, P_R, L_i, T_i)$. It sends $(\delta_{i1}, \delta_{i2}, U_i, \mathrm{PID}_i, P_R, L_i, T_i)$ to RSU_j through the OBU.

RSU_j receives $(\delta_{i1}, \delta_{i2}, U_i, \mathrm{PID}_i, P_R, L_i, T_i)$, and then, it computes $h_{i1} = h_2(\mathrm{ID}_j, \mathrm{PID}_i, U_i, T_i)$ and verifies whether the following equation holds.

$$\begin{aligned} &\delta_{i1}P = U_i + h_{i1}P_x, \\ &\because \delta_{i1}P = u_iP + h_{i1}xP = U_i + h_{i1}P_x. \end{aligned} \tag{2}$$

If (2) holds, RSU_j computes $\mathrm{RID}_i = \mathrm{PID}_i \oplus h_1(s_jL_i)$, $R_i = P_R - s_jl_iP = P_R - s_jL_i$, $h_{\mathrm{vi}} = h_0(RID_i, R_i)$, and $h_{i2} = h_3(\mathrm{ID}_j, \mathrm{PID}_i, P_R, L_i, T_i)$ and verifies whether the following equation holds.

$$\begin{aligned} \delta_{i2}P &= L_i + h_{i2}R_i + h_{i2}h_{\mathrm{vi}}P_{\mathrm{pub}}, \\ \because \delta_{i2}P &= l_iP + h_{i2}(r_iP + h_{\mathrm{vi}}sP), \\ &= L_i + h_{i2}R_i + h_{i2}h_{\mathrm{vi}}P_{\mathrm{pub}}, \end{aligned} \tag{3}$$

if both (2) and (3) hold, the vehicle is legal. RSU_j chooses a random number $u_j \in Z_q^*$ and computes $U_j = u_jP$, $\mathrm{sk} = h_2(\mathrm{ID}_j, \mathrm{RID}_i, u_jU_i, T_i)$, $h_j = h_4(\mathrm{ID}_j, \mathrm{RID}_i, U_j, \mathrm{sk}, T_i)$, and $\delta_j = h_js_j + u_j$. RSU_j sends $(\delta_j, U_j, \mathrm{PID}_i, \mathrm{ID}_j, R_j)$ to the vehicle V_i.

The vehicle V_i receives $(\delta_j, U_j, \mathrm{PID}_i, \mathrm{ID}_j, R_j)$, and then, it computes $h_{\mathrm{rj}} = h_0(\mathrm{ID}_j, R_j)$ and $h_j = h_4(\mathrm{ID}_j, \mathrm{RID}_i, U_j, \mathrm{sk}, T_i)$ and verifies whether the following equation holds.

$$\begin{aligned} \delta_jP &= h_jR_j + h_jh_{rj}P_{\mathrm{pub}} + U_j, \\ \because \delta_jP &= h_js_jP + u_jP = h_jP_j + U_j, \\ &= h_jR_j + h_jh_{rj}P_{\mathrm{pub}} + U_j. \end{aligned} \tag{4}$$

If (4) holds, the vehicle V_i computes $sk = h_2(\mathrm{ID}_j, \mathrm{RID}_i, u_iU_j, T_i)$, which is the session key between V_i and RSU_j.

The process of authentication and key agreement between vehicle and RSU is shown in Figure 2.

3.4. Message Signing Phase. When a vehicle needs to send a message M_i in the area covered by the roadside unit RSU_j, the TPD of the vehicle chooses a random number $v_i \in Z_q^*$ and the timestamp T_m and computes $V_i = v_iP$, $P_v = R_i + v_iP_j$, the pseudonym $\mathrm{PID}_i = \mathrm{RID}_i \oplus h_1(v_iP_j)$, and $\sigma_i = h_{\mathrm{rv}}h_{\mathrm{vi}}^{-1}(s_i + v_i) + h_{\mathrm{mi}}x$, where $h_{\mathrm{rv}} = h_2(M_i, \mathrm{PID}_i, R_{\mathrm{vi}}, T_m)$, $R_{\mathrm{vi}} = h_{\mathrm{vi}}^{-1}(R_i + V_i)$, and $h_{mi} = h_5(M_i, \mathrm{PID}_i, P_v, V_i, R_{\mathrm{vi}}, T_m)$. The TPD then broadcasts the signature $(\sigma_i, M_i, \mathrm{PID}_i, V_i, P_v, R_{\mathrm{vi}}, T_m)$.

3.5. Signature Verification. RSU_j receives $(\sigma_i, M_i, PID_i, V_i, P_v, R_{vi}, T_m)$, and then, it checks whether the timestamp T_m is within the valid time. If it is, RSU_j extracts the real identity of the vehicle $RID_i = PID_i \oplus h_1(s_jV_i)$and computes $R_i = P_v - s_jV_i$, $h_{vi} = h_0(RID_i, R_i)$, $R_{vi} = h_{vi}^{-1}(R_i + V_i)$, $h_{rv} = h_2(M_i, PID_i, R_{vi}, T_m)$, $h_{mi} = h_5(M_i, PID_i, P_v, V_i, R_{vi}, T_m)$, and verifies whether (5) holds.

$$\begin{aligned} h_{\mathrm{vi}}\sigma_iP &= h_{\mathrm{rv}}(R_i + V_i) + h_{\mathrm{rv}}h_{\mathrm{vi}}P_{\mathrm{pub}} + h_{\mathrm{vi}}h_{\mathrm{mi}}P_x, \\ \because h_{\mathrm{vi}}\sigma_iP &= h_{\mathrm{rv}}(s_i + v_i)P + h_{\mathrm{vi}}h_{\mathrm{mi}}xP, \\ &= h_{\mathrm{rv}}(R_i + h_{\mathrm{vi}}sP + V_i) + h_{\mathrm{vi}}h_{\mathrm{mi}}P_x \\ &= h_{\mathrm{rv}}(R_i + V_i) + h_{\mathrm{rv}}h_{\mathrm{vi}}P_{\mathrm{pub}} + h_{\mathrm{vi}}h_{\mathrm{mi}}P_x. \end{aligned} \tag{5}$$

If it holds, RSU_j accepts the message. If it does not, it means that the TPD of the vehicle is damaged. For example, suppose the attackers stole the private and group keys of the TPD, faked the identity RID_i', generated the pseudonym PID_i', forged the signature $(\sigma_i, M_i, \mathrm{PID}_i', V_i', P_v', R_{\mathrm{vi}}', T_m)$, and enabled it to satisfy (6). However, according to the TPD security assumption, this situation is extremely rare. If RSU_j detects that the TPD has been attacked, it immediately broadcasts that the signature $(\sigma_i, M_i, \mathrm{PID}_i', V_i', P_v', R_{\mathrm{vi}}', T_m)$ is not valid.

Other vehicles receive $(\sigma_i, M_i, \mathrm{PID}_i, V_i, P_v, R_{\mathrm{vi}}, T_m)$, and then, they check whether the timestamp T_m is within the valid time. If it is, the vehicles compute $h_{\mathrm{rv}} = h_2(M_i, \mathrm{PID}_i, R_{\mathrm{vi}}, T_m)$ and $h_{mi} = h_5(M_i, \mathrm{PID}_i, P_v, V_i, R_{\mathrm{vi}}, T_m)$ and verify whether the following equation holds.

$$\begin{aligned} \sigma_iP &= h_{\mathrm{rv}}R_{\mathrm{vi}} + h_{\mathrm{rv}}P_{\mathrm{pub}} + h_{\mathrm{mi}}P_x, \\ \because \sigma_iP &= h_{\mathrm{rv}}h_{\mathrm{vi}}^{-1}(s_i + v_i)P + h_{\mathrm{mi}}xP \\ &= h_{\mathrm{rv}}h_{\mathrm{vi}}^{-1}(R_i + h_{\mathrm{vi}}sP + V_i) + h_{\mathrm{mi}}P_x \\ &= h_{\mathrm{rv}}h_{\mathrm{vi}}^{-1}(R_i + V_i) + h_{\mathrm{rv}}P_{\mathrm{pub}} + h_{\mathrm{mi}}P_x \\ &= h_{\mathrm{rv}}R_{\mathrm{vi}} + h_{\mathrm{rv}}P_{\mathrm{pub}} + h_{\mathrm{mi}}P_x. \end{aligned} \tag{6}$$

If it does and the vehicles do not receive an invalid signature broadcasted by RSU_j within the specified time, the vehicles accept the message M_i.

3.6. Identity Extraction. When a valid message signature $(\sigma_i, M_i, \mathrm{PID}_i, V_i, P_v, R_{\mathrm{vi}}, T_m)$ is in dispute, it is necessary to track the real identity of a vehicle. RSU_j can extract the real identity of the vehicle through computing $\mathrm{RID}_i = \mathrm{PID}_i \oplus h_1(s_jV_i)$.

3.7. Updating the Group Key Phase. TA chooses a random number $w_i \in Z_q^*$ and the timestamp T_v and computes $W_i = xw_iP$, $\delta_t = sh_0(xw_iP, T_v) + xw_i$, $P_x = h_3(w_iP, xP, T_v)P$, where P_x is as a new group public key. TA broadcasts the signature $(\delta_t, W_i, T_v, P_x)$.

After the vehicles receive $(\delta_t, W_i, T_v, P_x)$, they compute $x^{-1}W_i$ and verify whether the following equation holds. If it does, the vehicles update the group key as $x = h_3(x^{-1}W_i, xP, T_v)$.

$$\delta_tP = h_0(W_i, T_v)P_{\mathrm{pub}} + xW_i. \tag{7}$$

4. Security Analysis

Under the random oracle model, the security model of [39] is used to prove the security of our scheme.

4.1. Proof of Safety

Lemma 1. *The authentication request message of the vehicle cannot be forged. When ECDLP is a difficult problem, our scheme can resist the forgery attack of adaptive chosen message.*

Proof. We assume that there is an attacker Ad who can successfully forge the request message of a vehicle in polynomial time ε. Given an ECDLP instance $(P, Q = xP, P, Q \in G, x \in Z_q^*)$, the challenger Ch can solve the ECDLP in polynomial time ε.

The challenger *Ch* sets system parameters paras $= \{E_p(a,b), p, q, G, P, P_{\text{pub}}, P_x, h_0, h_1, h_2, h_3, h_4, h_5\}$. Ch randomly chooses RID_i of a vehicle as the identity of the challenger Ch. Ch builds and maintains six hash lists: L_{hl}, where $l = 0, 1, 2, \ldots, 5$. Finally, Ch sends params to Ad.

h_1- *Oracle*. When Ad makes a query with θ, Ch checks whether the tuple (θ, τ_{h1}) is already in L_{h1} or not. If it is, Ch sends τ_{h1} to Ad. Otherwise, Ch randomly selects $\tau_{h1} \in Z_q^*$ and adds $(\theta, \tau_{h1})$to L_{h1}. Finally, Ch sends $\tau_{h1} = h_1(\theta)$ to Ad.

h_2- *Oracle*. When Ad makes a query with $(\text{ID}_j, \text{PID}_i, U_i, T_i)$, Ch checks whether the tuple $(\text{ID}_j, \text{PID}_i, U_i, T_i, \tau_{h2})$ is already in L_{h2} or not. If it is, Ch sends τ_{h2} to Ad. Otherwise, Ch randomly selects $\tau_{h2} \in Z_q^*$ and adds $(\text{ID}_j, \text{PID}_i, U_i, T_i, \tau_{h2})$to L_{h2}. Finally, Ch sends $\tau_{h2} = h_2(\text{ID}_j, \text{PID}_i, U_i, T_i)$ to Ad.

h_3- *Oracle*. When Ad makes a query with $(\text{ID}_j, \text{PID}_i, P_R, L_i, T_i)$, Ch checks whether the tuple $(\text{ID}_j, \text{PID}_i, P_R, L_i, T_i, \tau_{h3})$ is already in L_{h3} or not. If it is, Ch sends τ_{h3} to Ad. Otherwise, Ch randomly selects$\tau_{h3} \in Z_{q^*}$ and adds $(\text{ID}_j, \text{PID}_i, P_R, L_i, T_i, \tau_{h3})$ to L_{h3}. Finally, Ch sends $\tau_{h3} = h_3(\text{ID}_j, \text{PID}_i, P_R, L_i, T_i)$ to Ad.

Extract (RID_i). Ch builds and maintains the list $L_v = (\text{RID}_i, R_i, s_i)$. When Ad makes a query with RID_i and R_i, Ch checks whether the tuple (RID_i, R_i, s_i) is in L_v. If it is, Ch sends $s_i$to Ad. Otherwise, Ch randomly selects $s_i, h_{vi} \in Z_q^*$, lets $R_i = s_iP - h_{vi}P$, and adds them to L_v. Finally, Ch sends $L_v = (\text{RID}_i, R_i, s_i)$ to Ad.

Sign-Oracle. When Ad makes a query with (PID_i, T_i), Ch randomly selects $h_{i1}, h_{i2}, h_{vi}, \delta_{i1}, \delta_{i2} \in Z_q^*$ and sets $U_i = \delta_{i1}P - h_{i1}P_x$, $R_i = s_iP - h_{\text{vi}}P$, $L_i = \delta_{i2}P - h_{i2}R_i - h_{i2}h_{vi}P_{\text{pub}}$, and $P_R = R_i + s_jL_i$. Finally, Ch sends $(\delta_{i1}, \delta_{i2}, U_i, \text{PID}_i, P_R, L_i, T_i)$ to Ad.

Output. Finally, Ad outputs an authentication request message $(\delta_{i1}, \delta_{i2}, U_i, \text{PID}_i, P_R, L_i, T_i)$ with nonnegligible probability. According to the forgery lemma [40],.Ad chooses different h_{i1}' andh_{vi}' and generates another valid authentication request message $(\delta_{i1}', \delta_{i2}', U_i, \text{PID}_i, P_R, L_i, T_i)$ in polynomial time. At this time, the two authentication request messages satisfy the following:

$$\delta_{i1}P = U_i + h_{i1}P_x, \tag{8}$$

$$\delta_{i1}'P = U_i + h_{i1}'P_x, \tag{9}$$

$$\delta_{i2}P = h_{i2}R_i + h_{i2}h_{vi}P_{\text{pub}} + L_i, \tag{10}$$

$$\delta_{i2}'P = h_{i2}R_i + h_{i2}h_{vi}'P_{\text{pub}} + L_i. \tag{11}$$

From (8)–(11), we can obtain

$$(\delta_{i1} - \delta_{i1}')P = (h_{i1} - h_{i1}')P_x, \tag{12}$$

$$(\delta_{i2} - \delta_{i2}')P = (h_{i2}h_{vi} - h_{i2}h_{vi}')P_{\text{pub}}. \tag{13}$$

Now, according to (12) and (13), Ad outputs$x = (\delta_{i1} - \delta_{i1}')(h_{i1} - h_{i2}')^{-1}$, and $s = (\delta_{i2} - \delta_{i2}')(h_{i2}h_{vi} - h_{i2}h_{vi}')^{-1}$.However, solving x or s is an ECDLP problem. Furthermore, it is impossible for an adversary to solve the ECDLP problem in polynomial time. □

Lemma 2. *The authentication response message cannot be forged. Since ECDLP is difficult to solve, our scheme can resist the forgery attack of adaptive chosen message.*

Proof. We assume that there is an attacker Ad who can successfully forge an authentication response message in polynomial time. Given an ECDLP instance $(P, Q = xP, P, Q \in G, x \in Z_q^*)$, then the challenger Ch can solve the ECDLP with nonnegligible probability. The challenger Ch sets system parameters paras $= \{E_p(a,b), p, q, G, P, P_{\text{pub}}, P_x, h_0, h_1, h_2, h_3, h_4, h_5\}$. Ch builds and maintains six lists: L_{hl}, where $l = 0, 1, 2, \ldots, 5$. Finally, Ch sends params to Ad.

h_1-*Oracle*. When Ad makes a query with θ, Ch checks whether the tuple (θ, τ_{h1}) is already in L_{h1} or not. If it is, Ch sends τ_{h1} to Ad. Otherwise, Ch randomly selects $\tau_{h1} \in Z_q^*$ and adds (θ, τ_{h1}) to L_{h1}. Finally, Ch sends $\tau_{h1} = h_1(\theta)$ to Ad.

h_2-*Oracle*. When Ad makes a query with $(\text{ID}_j, \text{RID}_i, u_jU_i, T_i)$, Ch checks whether the tuple $(\text{ID}_j, \text{RID}_i, u_jU_i, T_i, \tau_{h2})$ is already in L_{h2} or not. If it is, Ch sends τ_{h2} to Ad. Otherwise, Ch randomly selects $\tau_{h2} \in Z_q^*$ and adds $(\text{ID}_j, \text{RID}_i, u_jU_i, T_i, \tau_{h2})$ to h_{h22}. Finally, Ch sends $\tau_{h2} = h_2(\text{ID}_j, \text{RID}_i, u_jU_i, T_i)$ to Ad.

Extract (ID$_j$). Ch builds and maintains the list $L_R = (ID_j, R_j, s_j)$. When Ad makes a query with ID_j, Ch checks whether the tuple (ID_j, R_j, s_j) is in L_{R}. If it is, Ch sends $s_j$to Ad. Otherwise, Ch randomly selects$s_j, h_{rj} \in Z_q^*$, lets $R_j = s_jP - h_{rj}P_{\text{pub}}$, and adds $(\text{ID}_j, R_j, s_j)$to L_{R}. Finally, Ch sends $L_R = (\text{ID}_j, R_j, s_j)$ to Ad.

Sign - Oracle. When Ad makes a query with (PID_i, T_i), Ch randomly chooses $sk, h_{r_j}, h_j, \delta_j \in Z_q^*$and sets$U_j = \delta_jP - h_jR_j - h_jh_{rj}P_{\text{pub}}$. Finally, Ch sends $(\delta_j, U_j, PID_i, ID_j, R_j)$ to Ad.

Output. Finally, Ad outputs an authentication request message $(\delta_j, U_j, PID_i, ID_j, R_j)$ with nonnegligible probability. According to the forgery lemma [40].Ad chooses different h_{rj}' and generates another valid authentication request message $(\delta_j', U_j, PID_i, ID_j, R_j)$ in polynomial time. Now, the two authentication request messages satisfy the following:

$$\delta_j P = h_j R_j + h_j h_{rj} P_{\text{pub}} + U_j, \tag{14}$$

$$\delta'_j P = h_j R_j + h_j h'_{rj} P_{\text{pub}} + U_j. \tag{15}$$

From (14) and (15), we can deduce the following expression:

$$\left(\delta_j - \delta'_j\right) P = h_j \left(h_{rj} - h'_{rj}\right) P_{\text{pub}}. \tag{16}$$

Next, Ch can output $s = (\delta_j - \delta'_j)(h_j(h_{rj} - h'_{rj}))^{-1} \bmod q$. However, solving s is an ECDLP, which is impossible for an adversary to solve in polynomial time. □

Theorem 1. *From Lemma 1 and Lemma 2, we know that when the ECDLP problem is difficult to solve, and the adversary cannot forge the authentication request message and response message, that is, our authentication scheme can resist adaptive chosen message forgery attack.*

Theorem 2. *The message signature cannot be forged. Since ECDLP is hard to solve, our scheme can resist the forgery attack of adaptive chosen message attack.*

Proof. We assume that there is an attacker Ad who can successfully forge an authentication response message in polynomial time. Given an ECDLP instance $(P, Q = xP, P, Q \in G, x \in Z_q^*)$, the challenger Ch can solve the ECDLP in polynomial time ε.

The challenger Ch sets system parameters paras $= \left\{E_p(a,b), p, q, G, P, P_{\text{pub}}, P_x, h_0, h_1, h_2, h_3, h_4, h_5\right\}$. Ch randomly chooses ID_j as the identity of the challenger Ch. Ch builds and maintains six lists: L_{hl}, where $l = 0, 1, 2, \ldots, 5$. Then, Ad adaptively queries the oracle machine to Ch, and Ch replies to Ad in the following way.

When Ad makes a query with (PID_i, M_i, T_m), Ch randomly chooses $h_{rv}, h_{mi}, \sigma_i \in Z_q^*$, and $V_i, P_v \in G$; furthermore, it sets $R_{\text{vi}} = h_{rv}^{-1}(\sigma_i P - h_{rv} P_{\text{pub}} - h_{mi} P_x)$. Finally, Ch sends $(\sigma_i, M_i, V_i, P_v, R_{vi}, T_m)$ to Ad.

Subsequently, Ad outputs a valid signature $(\sigma_i, M_i, V_i, P_v, R_{vi}, T_m)$ with a nonnegligible probability. According to the forgery lemma [40], Ad chooses different $h_{\text{mi}}{}'$ and generates another valid signature $(\sigma'_i, M_i, V_i, P_v, R_{vi}, T_m)$ in polynomial time. At this time, the two signatures satisfy the following relationships:

$$\sigma_i P = h_{\text{rv}} R_{\text{vi}} + h_{\text{rv}} P_{\text{pub}} + h_{\text{mi}} P_x, \tag{17}$$

$$\sigma'_i P = h_{\text{rv}} R_{\text{vi}} + h_{rv} P_{\text{pub}} + h'_{\text{mi}} P_x. \tag{18}$$

From (17) and (18), we can obtain the following equation:

$$(\sigma_i - \sigma'_i) P = (h_{mi} - h'_{mi}) P_x. \tag{19}$$

Now, according to (19), Ch can output $x = (\sigma_i - \sigma'_i)(h_{mi} - h'_{mi})^{-1} \bmod q$. However, solving for x is an ECDLP problem, which is impossible for an adversary to solve in polynomial time. Thus, our proposed signature scheme under the random oracle model is resistant against a chosen adaptive message attack. □

Theorem 3. *The key agreement of our scheme is secure under the ECCDH problem.*

Proof. Given an ECCDH instance, $Q_1 = x_1 P, Q_2 = x_2 P$, and $Q_3 = x_1 x_2 P$, where $x_1, x_2 \in Z_q^*$. In our key agreement, we let $Q_1 \longleftarrow U_i = u_i P$, $Q_2 \longleftarrow U_j = u_j P$, $Q_3 \longleftarrow u_i u_j P$. In this method, if the attacker Ad gets $u_i u_j P$ according to U_i, U_j, the key negotiated between the vehicle and RSU can be obtained. However, it is impossible for the adversary to solve the ECCDH problem in polynomial time, implying that the key agreement proposed in this study is secure. □

Theorem 4. *In the random oracle model, we can achieve conditional anonymity and traceability.*

Proof. In the proposed scheme, the authentication request message uses the pseudonym $\text{PID}_i = \text{RID}_i \oplus h_1(l_i P_j)$, where $L_i = l_i P$, $l_i \in Z_q^*$. According to ECDLP, it is not feasible for the adversary to solve $l_i P_j$ without knowing l_i. The request authentication signatures are $\delta_{i1} = u_i + h_{i1} x$ and $\delta_{i2} = h_{i2} s_i + l_i$, where $h_{i1} = h_2(\text{ID}_j, \text{PID}_i, U_i, T_i)$, $h_{i2} = h_3(\text{ID}_j, \text{PID}_i, P_R, L_i, T_i)$, and $u_i, l_i \in z_q^*$ are the random numbers. Every time a vehicle is certified, it can produce unrelated pseudonyms and different authentication requests. Similarly, the pseudonym is also used in message signature $\text{PID}_i = \text{RID}_i \oplus h(v_i P_j)$, $V_i = v_i P$. The message signature is $(\sigma_i, M_i, V_i, P_v, R_{vi}, T_m)$, $\sigma_i = h_{rv}(h_{vi}^{-1}(s_i + v_i)) + h_{mi} x$, $h_{\text{rv}} = h_2(M_i, \text{PID}_i, R_{\text{vi}}, T_m)$, and $h_{\text{mi}} = h_5(M_i, \text{PID}_i, P_v, V_i, R_{\text{vi}}, T_m)$. The pseudonym used in the signature is different every time. Therefore, the scheme can provide anonymity for vehicle users in authentication and message signature. In addition, this scheme can also realize the traceability of the real identity; RSU can calculate the real identity of the vehicle $\text{RID}_i = \text{PID}_i \oplus h_1(s_j L_i)$ through the private key. Similarly, through pseudonym of signature message $\text{PID}_i = \text{RID}_i \oplus h(v_i P_j)$, RSU can also calculate $\text{RID}_i = \text{PID}_i \oplus h_1(s_j V_i)$ using the private key. Therefore, this scheme can realize the traceability of identity. □

Theorem 5. *In the proposed scheme, we can achieve unlinkability.*

Proof. In our scheme, the authentication request message of the vehicle $(\delta_{i1}, \delta_{i2}, U_i, \text{PID}_i, P_R, L_i, T_i)$ is different for each session. Meanwhile, the signature message $(\sigma_i, M_i, \text{PID}_i, V_i, P_v, R_{\text{vi}}, T_m)$ is also different for each message. Therefore, all elements from the message of the vehicle are different, and any attacker cannot tell apart if two different messages from the same vehicle. Thus, our proposed scheme supports unlinkability. □

4.2. Other Security Analysis and Feature Comparison. From Theorem 1 and Theorem 2, it is ascertained that under the random oracle model, the authentication, key agreement, and message signature can resist adaptive chosen message forgery attacks. Additionally, there is no

need for the certification table or TA to participate in the certification between vehicle and RSU. Both authentication and message signature use timestamp, which can resist replay attack. In the authentication process, RSU first checks whether the group key signature is legal, and then, it verifies the vehicle private key signature. If the group key signature is illegal, the signature is discarded directly, which can resist DoS attack to a certain extent. In this scheme, the vehicle is equipped with a TPD, which stores the private key of the vehicle and the group key. Even if a single TPD is attacked, the attacker can only intercept the group key and the private key of the vehicle. The authentication, key agreement, and message signature all need the private key of the vehicle. Thus, the attacker can only forge the signature of a single vehicle, without affecting the communication security of other VANET nodes. The schemes [13, 24] keep the system key in the TPD of each vehicle; this requires a strong TPD security assumption. If a single TPD is successfully attacked, the whole system will not be secure. Table 2 provides the features comparison with other schemes. It can be seen from Table 2 that the proposed scheme has strong advantages in security and communication efficiency.

5. Performance Analysis

In this section, we analyze the computation cost and communication cost of message authentication.

5.1. Computation Performance Analysis. In this study, nonsingular elliptic curve cryptography is used, whereas bilinear pairing construction scheme is utilized in works [13, 24]. To compare at the same security level, we construct two 80 bit security level cryptographic operation schemes. Bilinear pairing cryptographic schemes are set as follows: e: $G_1 \times G_1 \longrightarrow G_2$. $\overline{E}$: $y^2 = x^3 + ax + b \bmod \overline{p}$ is a hyper singular curve with degree 2, where $\overline{p}$ is a 512 bit prime. G_1 is an additive group based on $\overline{E}$ with order $\overline{q}$ and $\overline{P}$ is the generator of G_1 with order $\overline{q}$. The elliptic curve cryptography of the same security level is set as follows: E: $y^2 = x^3 + ax + b \bmod p$ is a nonsuper singular elliptic curve, where p and q are 160 bit primes, $a, b \in Z_p^*$. G is an additive group on E. P is the generator of G with order q. Let T_{bp}, T_{bm}, and T_{ba} denote the execution time of bilinear pairing operation, scalar multiplication operation, and scalar addition operation, respectively. T_{em} and T_{ea} denote the execution time of scalar multiplication and scalar addition on elliptic curve cryptography, and T_H denotes the hash operation time of map-to-point. We use MIRACL cryptographic library, an i5-7200U processor with 2.5 GHz clock frequency and 8 GB memory in our experiment. The operating system is Windows 10. Table 3 provides the average execution time of cryptographic operations.

Next, we analyze the computation cost of the message signature and verification with the protocols given in Table 4. Message signature of LIAP [13] requires five bilinear scalar multiplication operations, one bilinear scalar addition operation, and one map-to-point operation; signature verification requires three bilinear pair operations, one bilinear scalar multiplication operation, and one map-to-point operation. Similarly, we can calculate the computation cost of message signature and signature verification for NECPPA [24], Wu et al.' scheme [6], and our scheme. As given in Table 4, the message signature cost of the vehicle is 2.475 ms in our scheme. Compared with LIAP and NECPPA, the message signature computation cost of our scheme is reduced by 74% and 87%, respectively. However, compared with Wu et al., it costs 1.65 ms more. Compared with LIAP and NECPPA, the cost of signature verification is reduced by 69% and 87%, and it is equal to Wu et al.'s scheme.

Figure 3 presents the comparisons of these computational costs graphically.

5.2. Communication Overhead. It can be seen from the analysis in the previous section that $\overline{p}$ is 64 bytes, G_1 is 128 bytes, and p is 20 bytes, G is 40 bytes. Suppose the timestamp is 4 bytes, the hash function value is 20 bytes, and the other nongroup elements have a value of 20 bytes. The signature message of the proposed method is $(\sigma_i, M_i, PID_i, V_i, P_v, R_{vi}, T_m)$, and the communication length is $20 + 20 + 20 + 40 + 40 + 40 + 4 = 184$ bytes. The signature message of LIAP is $(PID_i, M_s, PK_{Ri}, \sigma_i)$, and the communication length is $128 + 20 + 20 + 128 + 128 = 424$ bytes. The signature message of NECPPA is $(PID_i, \delta_i, M_i, ID_{RSU_j})$, and the communication length is $128 + 20 + 128 + 20 + 20 = 316$ bytes. The signature message of Wu et al.' scheme is $(m_i, PID_{vi}, T_i, T_{vi}, h_{ki}, R_i, \delta_i)$, and the communication length is $20 + 40 + 4 + 4 + 20 + 40 + 20 = 148$ bytes. Compared with LIAP and NECPPA, the proposed scheme can save 57% and 42% of the communication cost, respectively. Compared with Wu et al.' scheme, the communication length is slightly increased by 40 bytes. However, in the scheme proposed by Wu et al., RSU needs to store t pairs of the pseudonyms and local private keys (PID_{vi}, k_{vi}) [6] for each vehicle. When there are too many vehicles, it will cause a heavy burden on the memory of RSU. Similarly, each TPD also needs additional 60t bytes of storage space. The communication cost of message signature is provided in Table 5.

Figure 4 presents the comparisons graphically.

5.3. Comparison with Other Authentication Protocols. Wei et al.' protocols [37] use cosine similarity to realize the authentication for the intelligent and the authentication server. They have less computation cost and better accuracy compared with other implicit authentication schemes. The optimized computation complexity of two protocols is $3O(n^{2.3})$ and $3O(n^{2.3}) + 2Enc_p + Dec_p$, respectively, where n is the dimension of the multimodal behavior feature vector, and Enc_p and Dec_p are Pailler operations; our scheme is based on elliptic curve. Elliptic curve can achieve high security in 160-bit finite field. The complex operation used in our scheme is scalar multiplication operation. The complexity of scalar multiplication operation can be optimized to O (k), where k is the length of the coefficient, which is 160-bit in our scheme. In the process of mutual authentication

Table 2: Features comparison.

Feature	LIAP	NECPPA	Wu et al.	Our scheme
Using bilinear paring operation	Yes	Yes	No	No
Using ideal TPD	Yes	Yes	No	No
Requiring TA to participate in certification	Yes	Yes	Yes	No
Searching database	Yes	Yes	Yes	No

Table 3: Execution time of cryptographic operations.

Execution time	Value (ms)
T_{bp}	7.142
T_{bm}	1.445
T_{ba}	0.041
T_{em}	0.821
T_{ea}	0.006
T_H	2.228

Table 4: Computation cost of signature and verification of single message for various schemes.

Schemes	Message signature (ms)	Signature verification (ms)
LIAP	$5T_{bm} + 1T_{ba} + 1T_H \approx 9.494$	$3\,T_{bp} + 1\,T_{bm} + 1T_H \approx 25.099$
NECPPA	$4T_{bm} + 1T_{ba} + 1T_H \approx 8.049$	$3\,T_{bp} + 1\,T_{bm} + T_H \approx 25.099$
Wu et al.	$1T_{em} \approx 0.821$	$4T_{em} + 2T_{ea} \approx 3.296$
Our scheme	$3T_{em} + 2T_{ea} \approx 2.475$	$4T_{em} + 2T_{ea} \approx 3.296$

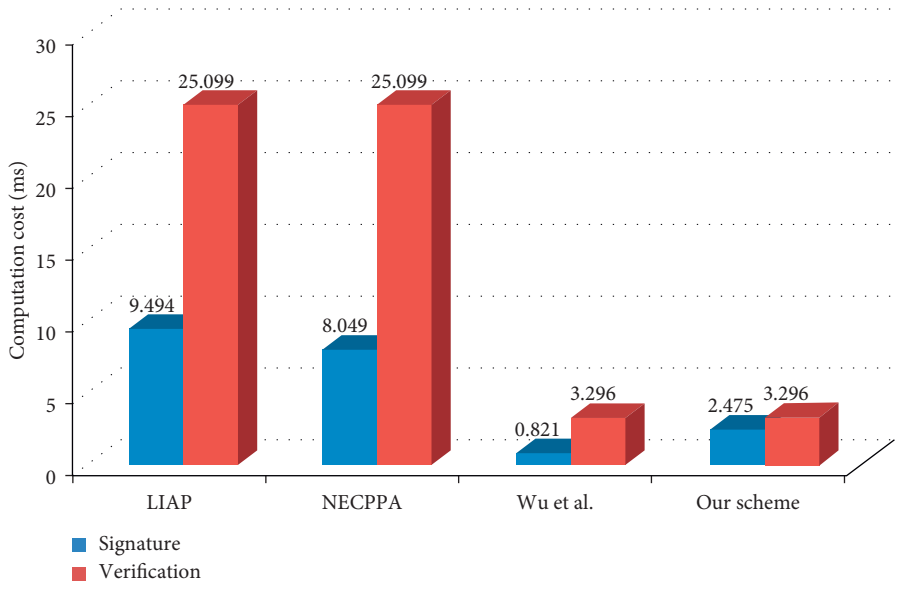


Figure 3: Comparison of computation costs.

between the vehicle and RSU, there are 15 scalar multiplication operations, and the computation complexity is 15O (160). It can be seen from the above analysis that when n is small, the work [37] has an advantage in computation cost, and when n is large, our scheme is better. In addition, in Wei et al.' protocols, the identity of the vehicle U_i is the same in different sessions, so they do not consider the unlinkability of sessions. In our scheme, we use different pseudonyms to realize unlinkability of the sessions.

Vinoth et at.' scheme [38] is a lightweight authentication and key agreement scheme, which is better than our scheme in terms of computation cost, communication cost, and storage cost. However, the scheme does not consider the internal attack. If one sensing device is attacked, the

Table 5: Communication cost of message signature in each scheme.

Schemes	Message signature	Communication overhead (bytes)
LIAP	$\mathrm{PID}_i, M_s, \mathrm{PK}_{\mathrm{Ri}}, \sigma_i$	424
NECPPA	$\mathrm{PID}_i, \delta_i, M_i, \mathrm{ID}_{\mathrm{RSU}_j}$	316
Wu et al.	$m_i, \mathrm{PID}_{\mathrm{vi}}, T_i, T_{\mathrm{vi}}, h_{\mathrm{ki}}, R_i, \delta_i$	148
Our scheme	$\sigma_i, M_i, \mathrm{PID}_i, V_i, P_v, R_{\mathrm{vi}}, T_m$	184

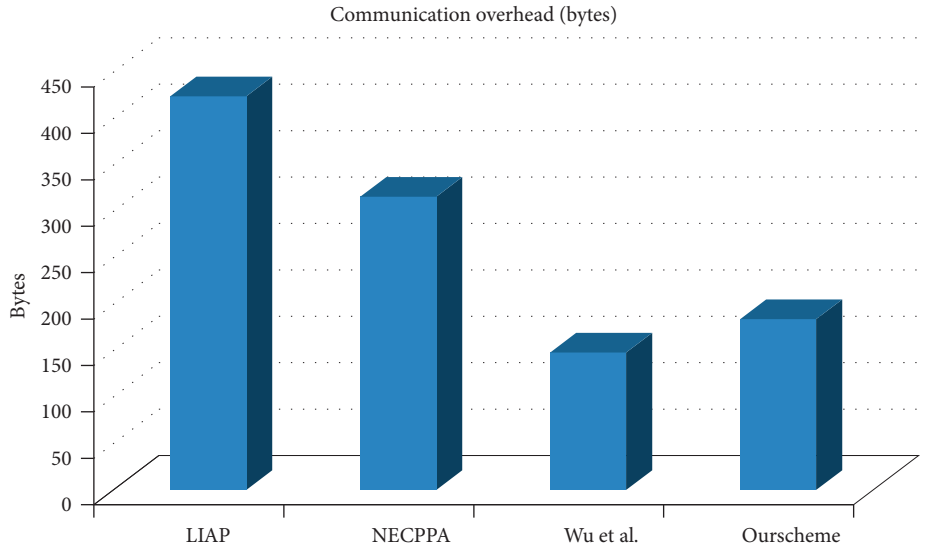


Figure 4: Comparison of communication overhead.

symmetric key $\mathrm{KEY}_{\mathrm{GWN\text{-}Ui}}$ and the session key SK can be obtained by the attacker. The attacker can monitor the communication between the user and the gateway node as well as between the user and other sensing devices. In our scheme, the vehicle is equipped with TPD, which stores the private key of the vehicle and the group key. Even if a single TPD is attacked, the attacker can only intercept the group key and the private key of the vehicle. The authentication, key agreement, and message signature all need the private key of the vehicle. Thus, the attacker can only forge the signature of a single vehicle, without affecting the communication security of other VANET nodes.

6. Conclusion

The instantaneous characteristic of VANET communication requires high efficiency in authentication and key agreement. Therefore, this study proposes an efficient anonymous authentication and key agreement scheme. The scheme includes mutual authentication and key agreement between vehicle and RSU, as well as signature and verification of the vehicle message. In the proposed scheme, an elliptic curve is used to improve the efficiency of computation and communication. Our authentication and key agreement scheme does not need to communicate with the third party authority or establish a local database, and furthermore, it avoids database query operation. It can effectively save the communication time and storage space of related nodes and is more suitable for VANET. Compared with other schemes, this scheme also has strong computing and communication advantages in message authentication. However, we do not address key negotiation and authentication between vehicles and vehicles. Lightweight and effective encryption methods to achieve anonymous authentication and communication between vehicles and vehicles is a worthy research direction. The implementation of anonymous authentication and key agreement based on channel condition is also one of the directions worthy of discussion [41, 42].

In this study, the authentication technology combined with cryptography is mainly presented. At present, deep learning and cloud computing are increasingly used in network applications. In the next step, more technologies such as deep learning [43, 44] and cloud computing can be combined into authentication and privacy protection of VANET.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Anhui University (KJ2018A0396,

KJ2017B015, KJ2019A0605, and KJ2020A0032), in part by the National Natural Science Foundation of China (61902140), and in part by the Anhui Provincial Natural Science Foundation (1908085QF288).

References

[1] J. Wu and Z. Cai, "Attribute weighting via differential evolution algorithm for attribute weighted naive bayes (WNB)," *Journal of Computational Information Systems*, vol. 7, no. 5, pp. 1672–1679, 2011.

[2] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[3] M. S. Kakkasageri and S. S. Manvi, "Information management in vehicular ad hoc networks: a review," *Journal of Network and Computer Applications*, vol. 39, pp. 334–350, 2014.

[4] M. Bayat, M. Barmshoory, M. Rahimi, and M. R. Aref, "A secure authentication scheme for VANETs with batch verification," *Wireless Networks*, vol. 21, no. 5, pp. 1733–1743, 2015.

[5] T. W. Chim, S. M. Yiu, L. C. K. Hui, V. O. K. Hui, and V. O. K. Li, "SPECS: secure and privacy enhancing communications schemes for VANETs," *Ad Hoc Networks*, vol. 9, no. 2, pp. 189–203, 2011.

[6] L. Wu, J. Fan, Y. Xie, J. Wang, and Q. Liu, "Efficient location-based conditional privacy-preserving authentication scheme for vehicle ad hoc networks," *International Journal of Distributed Sensor Networks*, vol. 13, no. 3, 2017.

[7] C. Cseh, "Architecture of the dedicated short-range communications(DSRC) protocol," in *Proceedings of the 1998 IEEE Conference on Vehicular Technology (VTC)*, May 1998.

[8] M. Raya and J. P. Hubaux, "Securing vehicular ad hoc networks," *Journal of Computer Security*, vol. 15, no. 1, pp. 39–68, 2007.

[9] B. Ying, D. Makrakis, and H. T. Mouftah, "Dynamic mix-zone for location privacy in vehicular networks," *IEEE Communications Letters*, vol. 17, no. 8, pp. 1524–1527, 2013.

[10] C. Zhang, X. Lin, R. Lu, and P. H. Ho, "An efficient RSU-aided message authentication scheme in vehicular communication networks," in *Proceedings of IEEE International Conference on Communications*, pp. 1451–1457, Phoenix, AZ, USA, 2008.

[11] R. Lu, X. Lin, H. Zhu, P. H. Ho, and X. Shen, "ECPP: efficient conditional privacy preservation protocol for secure vehicular communications," in *Proceedings of the 27th Conference on Computer Communications*, pp. 1229–1237, Phoenix, AZ, USA, April 2008.

[12] U. Rajput, F. Abbas, and H. Oh, "A hierarchical privacy preserving pseudonymous authentication protocol for VANET," *IEEE Access*, vol. 4, pp. 7770–7784, 2016.

[13] S. Wang and N. Yao, "LIAP: a local identity-based anonymous message authentication protocol in VANETs," *Computer Communications*, vol. 112, pp. 154–164, 2017.

[14] V. S. Miller, "Use of elliptic curves in cryptography," vol. 218, pp. 417–426, in *Proceedings of CRYPTO: Conference on the Theory and Application of Cryptographic Techniques*, vol. 218, pp. 417–426, Springer, Berlin, Germany, 1986.

[15] C. Zhang, R. Lu, X. Lin, P. H. Ho, and X. Shen, "An efficient identity-based batch verification scheme for vehicular sensor networks," in *Proceedings the 2008 IEEE Conference on Computer Communications*, pp. 246–250, Phoenix, AZ, USA, April 2008.

[16] J. L. Huang, L. Y. Yeh, and H. Y. Chien, "ABAKA: an anonymous batch Authenticated and key agreement scheme for value-added services in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 248–262, 2011.

[17] K. A. Shim, "${\cal CPAS}$: an efficient conditional privacy-preserving authentication scheme for vehicular sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1874–1883, 2012.

[18] C. C. Lee and Y. M. Lai, "Toward a secure batch verification with group testing for VANET," *Wireless Networks*, vol. 19, no. 6, pp. 1441–1449, 2013.

[19] H. Wang and Y. Zhang, "On the security of an anonymous batch authenticated and key agreement scheme for value-added services in VANETs," *Procedia Engineering*, vol. 29, no. 4, pp. 1735–1739, 2012.

[20] J. K. Liu, T. H. Yuen, M. H. Au, and W. Susilo, "Improvements on an authentication scheme for vehicular sensor networks," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2559–2564, 2014.

[21] M. Azees and P. Vijayakumar, "CEKD: computationally efficient key distribution scheme for vehicular ad-hoc networks," *Australian Journal of Basic and Applied Sciences*, vol. 10, no. 2, pp. 171–175, 2016.

[22] P. Vijayakumar, V. Chang, L. Jegatha Deborah, B. Balusamy, and P. G. Shynu, "Computationally efficient privacy preserving anonymous mutual and batch authentication schemes for vehicular ad hoc networks," *Future Generation Computer Systems*, vol. 78, pp. 943–955, 2018.

[23] M. A.P. Vijayakumar and L. J. Deboarh, "EAAP: efficient anonymous authentication with conditional privacy-preserving scheme for vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 9, pp. 2467–2476, 2017.

[24] S. M. Pournaghi, B. Zahednejad, M. Bayat, and Y. Farjami, "NECPPA: a novel and efficient conditional privacy-preserving authentication scheme for VANET," *Computer Networks*, vol. 134, pp. 78–92, 2018.

[25] I. Ali and F. Li, "An efficient conditional privacy-preserving authentication scheme for Vehicle-To-Infrastructure communication in VANETs," *Vehicular Communications*, vol. 22, 2019.

[26] J. Zhang, M. Xu, and L. Liu, "On the security of a secure batch verification with group testing for VANET," *International Journal of Network Security*, vol. 16, no. 5, pp. 355–362, 2014.

[27] M. A. Alazzawi, H. Lu, A. A. Yassin, and K. Chen, "Efficient conditional anonymity with message integrity and authentication in a vehicular ad-Hoc Network," *IEEE Access*, vol. 7, pp. 71424–71435, 2019.

[28] Y. Ming and X. Shen, "PCPA: a practical certificateless conditional privacy preserving authentication scheme for vehicular ad hoc networks," *Sensors*, vol. 18, no. 5, p. 1573, 2018.

[29] D. He, S. Zeadally, B. Xu, and X. Huang, "An efficient identity-based conditional privacy-preserving authentication scheme for vehicular ad hoc networks," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2681–2691, 2015.

[30] S. H. Islam, M. S. Obaidat, P. Vijayakumar, E. Abdulhay, F. Li, and M. K. C. Reddy, "A robust and efficient password-based conditional privacy preserving authentication and group-key agreement protocol for VANETs," *Future Generation Computer Systems*, vol. 84, pp. 216–227, 2018.

[31] J. Cui, J. Zhang, H. Zhong, and Y. Xu, "SPACF: a secure privacy-preserving authentication scheme for VANET with cuckoo filter," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10283–10295, 2017.

[32] H. Zhong, B. Huang, J. Cui, Y. Xu, and L. Liu, "Conditional privacy-preserving authentication using registration list in vehicular ad hoc networks," *IEEE Access*, vol. 6, pp. 2241–2250, 2017.

[33] X. Li, T. Liu, M. S. Obaidat, F. Wu, P. Vijayakumar, and N. Kumar, "A lightweight privacy-preserving authentication protocol for VANETs," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3547–3557, 2020.

[34] P. Vijayakumar, A. Bose, and A. Kannan, "Chinese remainder theorem based centralised group key management for secure multicast communication," *IET Information Security*, vol. 8, no. 3, pp. 179–187, 2014.

[35] J. Cui, X. Tao, J. Zhang, Y. Xu, and H. Zhong, "HCPA-GKA: a hash function-based conditional privacy-preserving authentication and group-key agreement scheme for VANETs," *Vehicular Communications*, vol. 14, pp. 15–25, 2018.

[36] J. Zhang, J. Cui, H. Zhong, Z. Chen, and L. Liu, "PA-CRT: Chinese remainder theorem based conditional privacy-preserving authentication scheme in vehicular ad-hoc networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 722–735, 2019.

[37] F. Wei, S. Zeadally, P. Vijayakumar, N. Kumar, and D. He, "An intelligent terminal based privacy-preserving multimodal implicit authentication protocol for Internet of connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[38] R. Vinoth, L. J. Deborah, P. Vijayakumar, and N. Kumar, "Secure multifactor Authenticated key agreement scheme for industrial IoT," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3801–3811, 2021.

[39] W.-B. Hsieh and J.-S. Leu, "An anonymous mobile user authentication protocol using self-certified public keys based on multi-server architectures," *The Journal of Supercomputing*, vol. 70, no. 1, pp. 133–148, 2014.

[40] D. Pointcheval and J. Stern, "Security proofs for signature schemes," *Advances in Cryptology*, vol. 4, pp. 387–398, 1996.

[41] K. Li, W. Ni, Y. Emami et al., "Design and implementation of secret key agreement for platoon-based vehicular cyber-physical systems," *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 2, 2019.

[42] K. Li, L. Lu, W. Ni, E. Tovar, and M. Guizani, "Secret key agreement for data dissemination in vehicular platoons," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 9060–9073, 2019.

[43] J. Wu, S. Pan, X. Zhu, and Z. Cai, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.

[44] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI 20)*, pp. 4981–4987, Yokohama, Japan, 2020.

Hindawi
Complexity
Volume 2021, Article ID 6661901, 10 pages
https://doi.org/10.1155/2021/6661901

Research Article

Multi-channel Convolutional Neural Network Feature Extraction for Session Based Recommendation

Zhenyan Ji,[1] **Mengdan Wu**,[1] **Yumin Feng**,[1] **and José Enrique Armendáriz Íñigo**[2]

[1]*School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China*
[2]*Department of Statistics, Computer Science and Mathematics, Public University of Navarre, Pamplona 31006, Spain*

Correspondence should be addressed to Zhenyan Ji; zhyji@bjtu.edu.cn

Received 29 November 2020; Revised 16 February 2021; Accepted 11 March 2021; Published 5 April 2021

Academic Editor: Dan Selisteanu

A session-based recommendation system is designed to predict the user's next click behavior based on an ongoing session. Existing session-based recommendation systems usually model a session into a sequence and extract sequence features through recurrent neural network. Although the performance is greatly improved, these procedures ignore the relationships between items that contain rich information. In order to obtain rich items embeddings, we propose a novel Recommendation Model based on Multi-channel Convolutional Neural Network for session-based recommendation, RMMCNN for brevity. Specifically, we capture items' internal features from three dimensions through multi-channel convolutional neural network firstly. Next, we merge the internal features with external features obtained by a GRU unit. Then, both internal features and external features are merged by an attention mechanism together as the input of the transformation function. Finally, the probability distribution is taken as the output after the softmax function. Experiments on various datasets show that our method's precision and recommendation performance are better than those of other state-of-the-art approaches.

1. Introduction

With the explosive growth of the information in the Internet era, recommendation systems have become an effective solution for users to deal with large amounts of information [1]. In order to have a better user experience, personalized recommendation systems have been applied to many scenarios, including movie recommendation [2, 3], music recommendation [4, 5], online shopping [6, 7], and other settings. In the recommendation scenario, the user behavior is modelled as a session. The session consists of the sequence of clicks performed by the user on items. The first time the user clicks on an item is regarded as the beginning of a session, and the last item of the user's continuous click browsing is the end of the session. Thus, the session contains the time series of user behavior and information between users and items [8, 9].

Traditional recommendation systems are mainly divided into recommendation systems based on collaborative filtering (CF), content-based recommendation systems (CB), and hybrid recommendation systems (HRS) [10]. CF-based recommendation systems build user preference models through the similarity of users or/and items. In addition, the CB recommendation systems state recommendations based on the content of item characteristics [11]. The former does not require contextual features; it only needs to train the matrix factorization model. The latter has good interpretability. In order to combine the advantages of both, HRS emerge to extract information from item attributes [12], users' social networks [13], and item comments [14].

On the other hand, in recent years, deep learning technology has been widely used in recommendation systems [15]. At the same time, powerful cloud computing capabilities have also laid the cornerstone for the development of deep learning [16]. For example, edge computing technology has made it possible to use machine learning technology to achieve intelligent network optimization [17]. Among many neural models, the recurrent neural network [18] approach was the first to be used. Afterwards, the community took into account the rich features of data.

Hence, the user temporal behavior is used in data augmentation [19]. Recently, STAMP [20] and SG-RNN [21] apply graph neural network to capture users' long-term and short-term interests as global interests and the last time the user clicks on the item as the current interest for recommendation.

Although the aforementioned methods achieve greater improvements, they still have some limitations. Firstly, a large number of session recommendation systems are based on users' historical behavior information. Without a large amount of user information, these recommendation systems are not be able to make proper recommendations. Secondly, the sequential features thanks to time stamp are fully captured, but the information between items is ignored.

To overcome the limitations mentioned above, we propose a Recommendation Model based on Mutichannel Convolutional Neural Network (RMMCNN). The main contributions are as follows:

(i) We introduce a multichannel convolutional neural network to extract item information in the context of a session.

(ii) To embed richer features, we use graph neural network to extract sequence features and internal features and then combine them as the final embedding vector representation through an adaptive mechanism.

(iii) Experiments are performed to compare our model with the baseline models. The results indicate that Precision and Mean Average Precision have been increased by at least 0.37% and 0.52%, respectively.

2. Related Work

Conventional recommendation methods include CF, CB, and HRS systems. In recent years, neural networks have greatly improved the performance of recommendation systems, including recurrent neural network, convolutional neural network, and graph convolutional neural network. Recurrent neural network [21–23] can extract users' historical click sequence features. Convolutional neural network [24, 25] can extract different local features of items and generate the corresponding item vector. Graph neural network [21, 26, 27] can learn graph structure data and capture vector embeddings of different nodes. These characteristics enable the neural network to learn more features. Therefore, the neural network can achieve better recommendation performance than conventional recommendation methods.

2.1. Conventional Methods

2.1.1. Collaborative Filtering (CF). Sarwar et al. [28] consider the impact of items on recommendation performance. This work analyzes the user-item matrix to identify different relational items and then uses relational items for indirect calculations. Cheng et al. [29] propose a collaborative filtering method based on user interest sequences. They introduce the similarity of users in the sequence dimension and extract the length of the user's longest common subinterest sequence and the total number of users. The number of common subinterest sequences is used to extract the information hidden in the sequence.

2.1.2. Content Based Recommendation (CB). Putri et al. [30] use the supervised learning method to represent and learn the data of 3700 articles in a vector space and apply a K-Neighbor algorithm for metrics. In order to alleviate the cold start problem, Zhang et al. [31] build the learned feature relationship matrix to extract user preference information hidden in content features. Trinh et al. [32] construct an association matrix between events and the user characteristics content. Concurrently to this, they also combine temporal and spatial relationships together with user interests to make recommendations to friends of key users.

2.1.3. Hybrid Recommendation. Hybrid recommendation algorithms aim to inherit the advantages of CF and CB recommendation algorithms. Rojsattarat and Soonthornphisaj [33] improve the recommendation performance using support vector machines, which map the information to the Euclidean space in order to extract features. Kiewra [34] utilizes the similarity between search and recommendation and uses positive and negative feedback to enhance the range of recommendations. Kolahkaj et al. [35] give importance to certain features of data such as time, location, user's hidden rating, and geographic information location. Then, it combines this information with collaborative filtering, context awareness, and other methods to make recommendations dynamically.

2.2. Deep Learning Methods

2.2.1. Recurrent Neural Network (RNN). Based on the RNN [22], the user's next clicked item can be predicted through similarity. Taking into account the essential characteristics of the item sequence, Long Short-Term Memory (LSTM) is used to capture the similarity between sequences. Xia et al. [23] combine RNN and an attention mechanism to learn about the session, sequence characteristics, and session context information. This fully mines the sequence characteristics of user sessions through recurrent neural network. Wu et al. [21] use the GRU unit to capture the sequence features and combine them with the last item in a session. Then, they are both generated into potential vectors for recommendation.

2.2.2. Convolutional Neural Network (CNN). Cai et al. [24] propose a multi-domain recommendation method based on CNN. It uses the generated user and item preference vectors to predict product ratings through a decomposition machine. Gao et al. [25] establish the CNN to capture the user's sequential features as positive feedback information and obtain negative feedback information through confrontation training, respectively. Afterwards, it combines both

feedbacks to generate action value functions for recommendation.

2.2.3. Graph Neural Network (GNN). GNNs are excellent in node information and graph structure information extraction. Fan et al. [26] propose a GNN framework for user-item graphs and their interactions to model two graphs and heterogeneous intensities. Xian et al. [27] construct the framework that combines the GNN with the repetitive exploration mechanism. It dynamically processes the sequence in a session through the graph structure and captures the complexity between items through the graph neural network. Wu et al. [21] use GNN and the GRU unit to generate a latent vector representation of a session sequence information and apply an attention mechanism to combine global and local user preferences.

3. The Proposed Model

In this section, we firstly present the proposed RMMCNN model. Then, we formulate the problem and introduce the process of our proposed model. Our model includes five steps: (1) knowledge distillation, (2) external feature extraction, (3) internal feature extraction, (4) joint feature extraction, and (5) possibility prediction.

3.1. RMMCNN Framework. Figure 1 introduces the framework of the proposed RMMCNN method. At first, all sessions are fed into a vector space via a directed graph. The items clicked by the user are nodes of the knowledge graph. The direction and connections of nodes in the graph represent the sequence of users that clicked two adjacent items consecutively. If we take into consideration the fact that some users may have clicked the same items in the very same sequence order, we have normalized each edge. We learn the latent vector representation of items through a graph neural network, so that each session will generate the corresponding embedding vector. After the embedding vector generation through the knowledge graph, the embedding vector containing the sequence features is generated by means of the GRU unit. More precisely, the session node embedding vector is propagated among different nodes through the GRU, not only extracting the features of neighboring nodes but also combining these neighboring features as the input of the graph neural network. Following that, the reset gate in the GRU determines whether the information should be kept or dropped, whereas the update gate refreshes all nodes to ensure the convergence of the results. Then, we extract external features and internal features separately for these embedding vectors. Internal features are extracted through a multi-channel convolutional neural network, which has three channels, where each channel has its own weight parameter, and we extract separately the internal features of different dimensions. The results of all channels will be combined through the attention mechanism as the final internal embedding. Finally, the node (resp., item) click probability is generated through linear and softmax transformations.

3.2. Problem Formulation. The goal of a session-based recommendation system is to predict for each session which item is going to be clicked next. Let $V = \{v_1, v_2, v_3, \ldots, v_n\}$ denote the set of all unique items in the sessions. All the sessions are composed of a series of items. Let $s = [v_1^s, v_2^s, v_3^s, \ldots, v_m^s]$ represent a history session sorted by time, where $v_i^s \in V$ represents that the user clicked the item i in session s. In this session, our goal is to predict where the user is going to click next, that is, v_{m+1}^s, in the context of the session.

Next, we are going to describe the steps that we follow to obtain the next clicked item in a session.

3.3. Knowledge Distillation. Recall that, within each session s, each item v is represented as a node in the directed graph $G_s = (V_s, \xi_s)$. The direction of each edge between nodes represents that the user clicked both items consecutively, according to the direction of the edge. Thus, a user in a session s firstly clicks item $v_{m-1}^s \in V$ and, immediately after that, clicks the item $v_m^s \in V$, which we denote as $(v_{m-1}^s, v_m^s) \in \xi_s$. In the directed graph, each item would be embedded into a unified embedding space. Let $N = \{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \ldots, \mathbf{n}_n\}$ denote the item embedding vectors, where $\mathbf{n} \in \mathfrak{R}^{\mathbf{d}}$ indicates the vector embedding of item v and d indicates its dimensionality, respectively. Let $s = [\mathbf{n}_1^s, \mathbf{n}_2^s, \mathbf{n}_3^s, \ldots, \mathbf{n}_m^s]$ denote the session s in the graph G_s^u. For this session s, the proposed RMMCNN will output probabilities $\widehat{y}$ for all the items, where $\widehat{y} = \{\widehat{y}_1, \widehat{y}_2, \widehat{y}_3, \ldots, \widehat{y}_n\}$. The top-K items in $\widehat{y}$ will be chosen as the candidate items.

3.4. External Feature Extraction. The session contains the users' clicked items within a period of time. In order to predict which item the user will click and the user behavior, we need to extract the representations of the session. At the same time, users' interests will change over time, and the temporal characteristics of the session should also be extracted. Therefore, we use the GRU joint attention mechanism to represent these external features.

GRU controls the flow of information through gates. GRU uses two gates, combining the input and forget gate of the LSTM into the update gate. The update gate determines the ratio of the previous value and the current one. The computation formula of the update gate is as follows:

$$\mathbf{z}_t^s = \psi\left(\mathbf{W}_{nz}\mathbf{n}_t^s + \mathbf{W}_{hz}\mathbf{n}_{t-1}^s\right), \tag{1}$$

where $\psi(\cdot)$ denotes the sigmoid function:

$$\psi(x) = \frac{1}{1 + e^{-x}}. \tag{2}$$

The reset gate establishes whether the current candidate state needs to depend on the network previous state and the weight of this dependency:

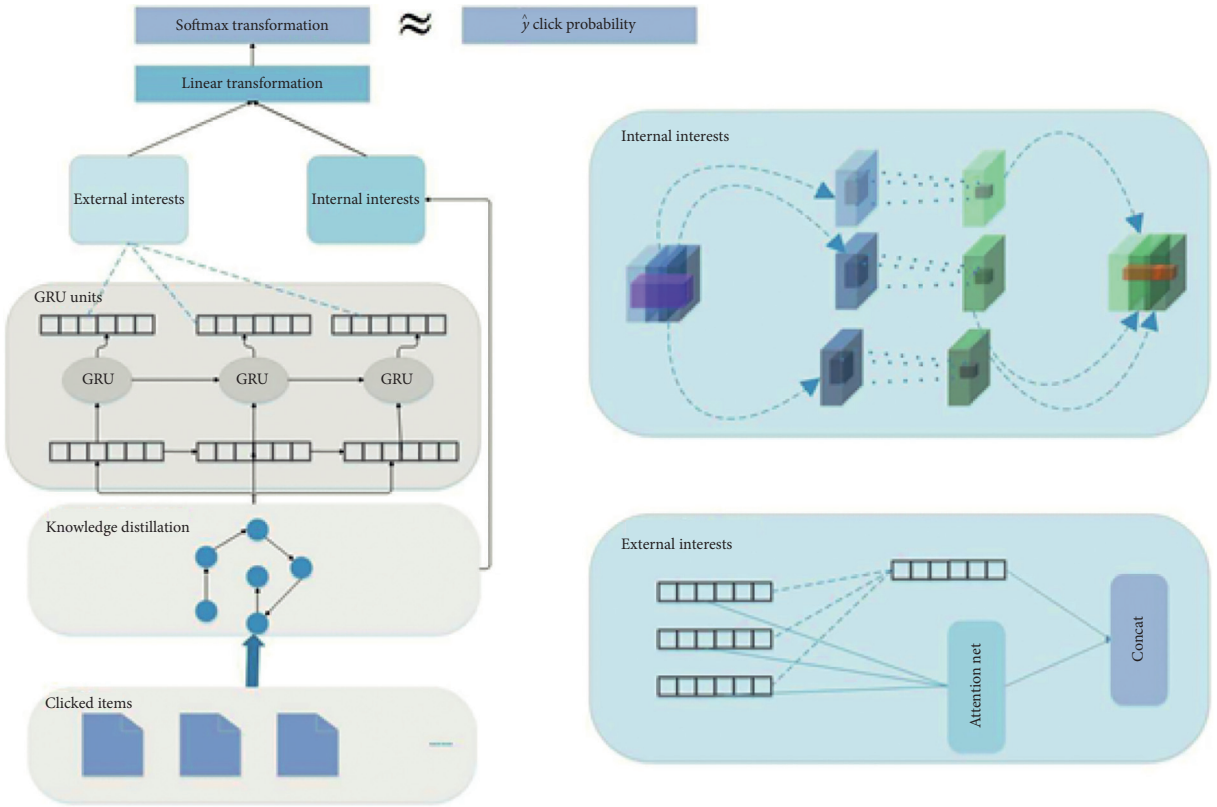


Figure 1: Illustration of the RMMCNN framework.

$$\mathbf{r}_t^s = \psi\left(\mathbf{W}_{nr}\mathbf{n}_t^s + \mathbf{W}_{hr}\mathbf{n}_{t-1}^s\right). \tag{3}$$

We need to estimate another intermediate value, which is the memory value. It is defined as

$$\mathbf{c}_t = \tanh\left(\mathbf{W}_{mn}\mathbf{n}_t^s + \mathbf{W}_{rc}\left(AGG\{(\mathbf{h}_{t-1}\mathbf{r}_t^s)\}\right)\right), \tag{4}$$

which is determined by the memory value in the previous state and the current input value. AGG is the aggregator function. We have chosen to use element-wise multiplication of the vectors. Thus, the state value of the hidden layer is defined as

$$\mathbf{h}_t = AGG\left(\{(1-\mathbf{z}_t^s)\mathbf{c}_t\}\right) + AGG\left(\{\mathbf{z}_t^s\mathbf{h}_{t-1}\}\right), \tag{5}$$

which is the weighted combination of the memory value of the current moment and the previous state value. It is important to note that, in the above formulae, $\mathbf{W}_{**}$ is the corresponding weight matrix.

After the GRU extracts all the features, the session s can be denoted as $\mathbf{v}_t^s = [\mathbf{v}_1^s, \mathbf{v}_2^s, \mathbf{v}_3^s, \ldots, \mathbf{v}_m^s]$.

3.5. Internal Feature Extraction. Following the notations used in Section 3, we use $N = \{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \ldots, \mathbf{n}_n\}$ to denote the raw input items sequence. We use a multi-channel convolutional neural network to extract the rich internal item features. Besides, we feed $\mathbf{v}_t$ into the neural network for further processing.

We expand the original two-dimension embedding vector into a higher dimension vector by firstly performing a convolution operation, which considers the difference of the internal features extraction with different convolution kernels. Our approach is loosely based on the RGB image processing, and we capture the different dimension features with a multi-channel convolutional neural network that we use to build the future maps. We set the convolution kernel to [1, 1, 1, 1], [1, 1, 1, 2], and [1, 3, 1, 1], respectively. Thus, the three convolution channel results can be expressed as $\widetilde{v}_R^t$, $\widetilde{v}_G^t$, and $\widetilde{v}_B^t$.

Then we merge these future maps with a linear transformation.

$$s_c = f\left(\widetilde{v}_R^t, \widetilde{v}_G^t, \widetilde{v}_B^t\right), \tag{6}$$

where $f(\cdot)$ is a linear transformation function. Once we extract the features, we perform a dimension reduction for subsequent processing to obtain the vector $\widetilde{s}_c$.

On the other hand, we extract the last item as v_l^s, which is denoted as v_m^s; that is, $v_l^s = v_m^s$.

3.6. Joint Feature Extraction. In order to maximize the representation of information, we aggregate the external and internal features together. Firstly, we aggregate all node embedding vectors and the last item:

$$\widetilde{s}_e = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i, \tag{7}$$

where $\alpha_i = \mathbf{v}^T\sigma(\omega_1\mathbf{v}_l^s + \omega_2\mathbf{v}_i^s + b)$, $i \in [1, m]$, and $\mathbf{v} \in \Re^{d_1}$ controls the weights of item vectors.

Then, we aggregate the result into the final feature. After this, we compute the hybrid embedding vector $\widetilde{s}_f$ through function transformation over the combination of the last clicked item and the external features:

$$\widetilde{s}_f = \boldsymbol{\beta}[\widetilde{s}_e; \widetilde{s}_c], \tag{8}$$

where matrix β compresses two combined embedding vectors into the latent space $\mathfrak{R}^d$.

3.7. Probability Prediction. After obtaining the representation of the session, we calculate the score $\widehat{g}_i^s$ as follows:

$$\widehat{g}_i^s = \widetilde{s}_f^T \boldsymbol{\omega}_3 v_i^s, \tag{9}$$

where ω_3 is the corresponding conversion dimension matrix and $\widehat{g}_i^s \in \mathfrak{R}^n$ represents the similarity scores with regard to each candidate item.

The score is transformed through the softmax function in the following:

$$\widehat{y}_i = \text{softmax}(\widehat{g}_i^s), \tag{10}$$

where $\widehat{y} \in \mathfrak{R}^n$ denotes the next clicked item probability for a given user in the context of the given session.

In our work, we adopt the cross-entropy as the loss function, and it is defined as follows:

$$\text{Loss}(\widehat{y}) = -\sum_{i=1}^{n} y_i \log(\widehat{\mathbf{y}}_i) + (1 - \mathbf{y}_i)\log(1 - \widehat{\mathbf{y}}_i), \tag{11}$$

where $\mathbf{y}$ denotes the user actual clicked items in the session.

The recommendation model can be trained through the backpropagation algorithm, and then the parameters in the model can be updated. In this process, we use the Adam optimizer [36] to train the parameters in the RMMCNN model.

4. Experiments and Analysis

In this part, we conduct three groups of experiments on two real world datasets. The datasets come from the RecSys 2015 Challenge called Yoochoose and CIKM Cup 2016 Challenge called Diginetica. The first experimental setup compares the recommendation performance of different models, whereas second group experiment compares the recommendation performance of different session embedding methods. Finally, the third group compares the recommendation performance of different evaluation criteria. Our experiments are based on TensorFlow 1.4.0 and Python 3.6.

4.1. Experiment Settings

4.1.1. Datasets. Yoochoose contains a series of click events of users on e-commerce websites, and these click events can be used to predict whether the user intends to click a certain product. Diginetica contains a large amount of information such as searches, logs, product data, and transaction data. In this paper, we only use transaction data.

Considering the existence of some noisy data in the sessions, we will filter out those values whose session length is 1 and those items that have been clicked fewer than 5 times, following [20, 21]. Then, we separate the two datasets and divide them into a training dataset and a test dataset, respectively. The reader is referred to Table 1 for a more detailed description. Considering that the Yoochoose dataset is quite large, we sort the sequences in the Yoochoose dataset and obtain the latest fractions 1/64 and 1/4 of the entire sequence according to time. We refer to them as Yoochoose 1/4 and Yoochoose 1/64.

4.1.2. Data Availability Statement. There are two datasets used in this paper: one is Yoochoose and the other is Diginetica. The Yoochoose dataset comes from the 2015 ACM RecSys Challenge. The content is a series of click events performed by users during a typical session in an e-commerce website. The data files include training data files and test data files. The former contains click events and purchase events, and the latter contains files. Each click event contains the session ID, timestamp, item ID, and item category. The Yoochoose dataset can be accessed at https://2015.recsyschallenge.com/challenge.html. The Diginetica dataset comes from CIKM Cup 2016. The dataset contains product data and transaction data. In this paper, we only use transaction data. The Diginetica dataset can be obtained from the following link: https://competitions.codalab.org/competitions/11161#learn_the_details-data2.

4.2. Baselines. In order to measure the performance of the proposed model, we compare the model with the following baseline algorithms.

(i) **NARM** [37] uses an attention mechanism to obtain the features in the hidden state to enhance the original information, which emphasizes the main purpose of the user in the current session. It proposes a neural attention recommender to solve the problem of lack of user purpose analysis in a session-based recommendation setting. NARM proposed a hybrid encoder to simulate the user's sequential behavior, capture the user's main purpose in the conversation, and merge this information as the final user behavior information representation.

(ii) **STAMP** [20] combines the current session information with the last clicked item in the current session. This mainly solves the problem of user behavior prediction based on anonymous sessions. It considers the impact of the user's current operation on the next clicked item as a tradeoff with the short-term memory model. It combines the short-term attention model with the original long-term memory model to extract the current and long-term user interest and generates the final interest of the user.

(iii) **SRGNN** [21] uses graph neural networks and GRU units to generate node latent vector representations.

Table 1: The statistics of the datasets.

Datasets	Yoochoose 1/4	Yoochoose 1/64	Diginetica
No. of clicks	7,980,529	565,552	982,961
No. of items	30,660	17,694	43,097
No. of train sessions	5,917,746	369,859	719,470
No. of test sessions	55,898	55,898	60,858
Average length (no. of clicks/session)	5.71	6.16	5.12

This approach tries to solve the problem of accurate user vector generation. SRGNN models user behavior as a graph structure data and captures item conversion information with a graph neural network. Then, the final embedding vector is generated through a linear transformation and the recommendations are made based on user's clicked items sequence and the last clicked item in the session.

4.3. Evaluation Metrics. We use often-used Precision (P) and Mean Reciprocal Rank (MRR) evaluation metrics to evaluate the performance of the RMMCNN model.

P@20: P@K measures predictive accuracy in recommendation systems. P@K describes the ranking ratio of recommended items accuracy in the recommendation lists, and it is defined as follows:

$$P@K = \frac{c_{\text{hit}}}{|C|}, \tag{12}$$

where $|C|$ denotes the total count of test data and c_{hit} represents the count of the hit data in the top-K ranking list.

MRR@20: MRR@K measures the accuracy of recommended clicked items, ordered by the probability of correctness. Given $K = 20$, if the right clicked item is suggested in apposition greater than 20, it will be set to zero:

$$MRR@K = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{1}{\text{rank}_i}, \tag{13}$$

where rank_i denotes users' first item ranking position in the recommendation list.

4.4. Parameter Settings. We set the latent vectors' dimensionality to $d = 100$ for the two datasets, like the settings of [20, 21, 37]. Besides, all parameters are set initially by a Gaussian distribution $N(0, 0.1^2)$, where its mean is 0 and $1/\sqrt{d}$ is its standard deviation. The initial learning rate is set to 0.001 and the batch size is set to 100. Recall that we use Adam optimizer [36] to update all parameters. For NARM, we set the mini-batch size to 512, the learning rate to 0.001, and the epochs to 30, respectively. For STAMP, the learning rate decay is set to 1 and the latent vectors' dimensionality is set to 100. For SRGNN, we set the learning rate decay to 0.1 behind every 3 epochs and the dimensionality is also 100. These key variables are shown in Table 2.

Table 2: The key variables of RMMCNN.

Variables	Value
Dimension	100
Learning rate	0.001
Batch size	100
Hidden size	100
Output size	100
L2 penalty	10^{-5}

4.5. Experimental Results. In this section, we compare RMMCNN in various aspects. Firstly, we compare RMMCNN with state-of-the-art methods. Then, we compare the performance of RMMCNN methods with variants of session embeddings. Finally, we compare the performance of RMMCNN with different evaluation metrics.

4.5.1. Comparison with Baseline Methods. To compare the performance of our proposed model, we compare it with some baseline models: NARM, STAMP, and SRGNN. The results show that RMMCNN outperforms all of them. The overall output is given in Figures 2 and 3; further detailed overview of the results can be seen in Table 3. RMMCNN aggregates the external and internal features of the session into the final data and makes recommendations based on both data. Compared with the current mainstream recommendation methods based on neural networks, RMMCNN shows excellent performance. NARM captures the user's overall interest through a cyclic neural network, and STAMP adds the last clicked event to the information extraction of the recommended item to achieve the purpose of information enhancement. These neural network-based methods achieve better recommendation performance than other methods. Figure 2 shows that STAMP is very low on the Diginetica dataset; this may be due to the fact that STAMP only uses the transition between users' last click item and users' historical click item. This information may not be enough to predict user session behavior. Therefore, the performance of STAMP in Diginetica dataset is very poor. SRGNN [21] generally considers the sequence characteristics of the user's click items and the user's last click event, which also combines the global characteristics and the local characteristics of the last click event for recommendation.

Our proposed model, RMMCNN, considers not only the external characteristics but also the internal characteristics of the session. Therefore, it combines the sequential characteristics of the session, the internal item characteristics of the session, and the last clicked event of the user to make recommendations. Simultaneously, RMMCNN uses an attention mechanism to automatically learn the weight

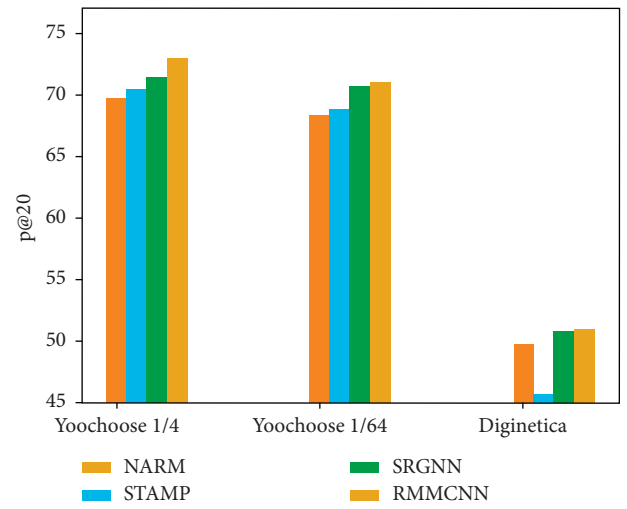


Figure 2: P@20 of different models.

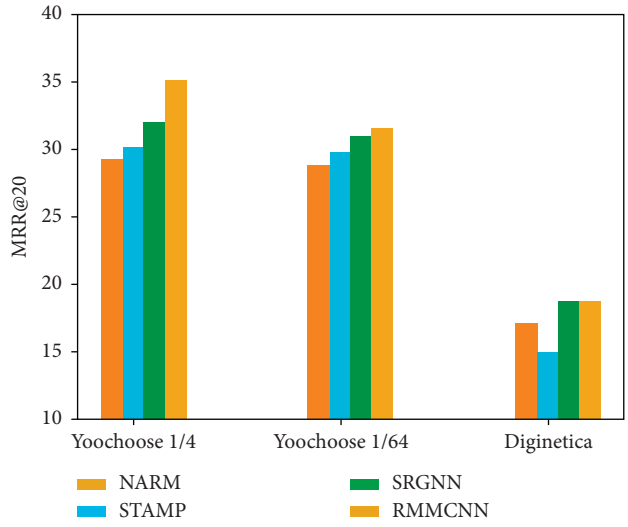


Figure 3: MRR@20 of different models.

Table 3: The performance of different datasets (K = 20).

Datasets	Yoochoose 1/4 (%)		Yoochoose 1/64 (%)		Diginetica (%)	
Measures	P@K	MRR@K	P@K	MRR@K	P@K	MRR@K
NARM	69.73	29.23	68.32	28.63	49.70	16.17
STAMP	70.44	30.00	68.74	29.67	45.64	14.32
SR-GNN	71.36	31.89	70.57	30.94	50.73	17.59
RMMCNN	72.93	35.04	70.94	31.46	51.82	18.19

representation. Thus, RMMCNN has obtained a richer user session representation and can make better recommendations for different user behaviors.

4.5.2. Comparison with Variants of Session Embeddings. We conduct distributed experiments on the model to verify the rationality of the RMMCNN model connection method

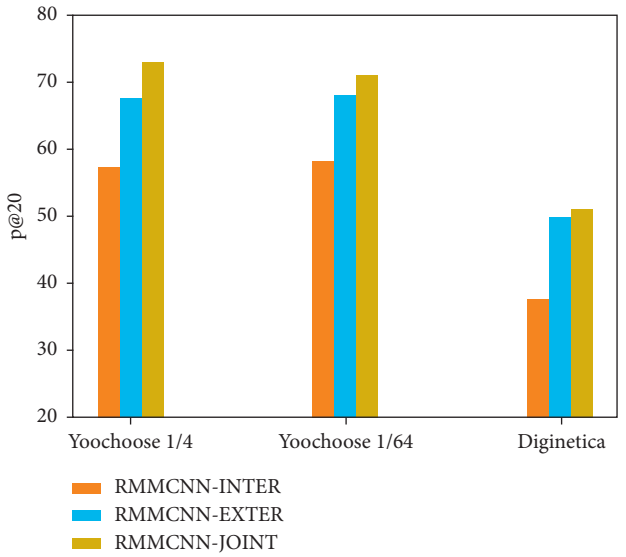


Figure 4: P@20 of different connection schemes.

and conduct experiments on the internal and external recommendation methods. We define the sequence feature of the session as an external feature and the relationship between users and items as an internal feature. We call the two methods RMMCNN-EXTER and RMMCNN-INTER, respectively.

(i) RMMCNN-EXTER: We first model the session as a directed graph and then extract the sequence features of the session through the GRU unit. Finally, we obtain different weight representations through the attention network.

(ii) RMMCNN-INTER: We obtain the embedding vector of the session to represent the internal relationship through a multichannel convolutional neural network. After that, we can get richer embedded vector information.

(iii) RMMCNN-JOINT: We combine the external vector with the internal vector of the session to generate the final session embedding representation.

Through Figures 4 and 5, we can see that the recommendation performance of RMMCNN-JOINT is better than that of the other two recommendation models. This is because RMMCNN-JOINT embeds more user information, and the content representation for recommendation is also richer. More details can be viewed in Table 4.

4.5.3. Comparison with Different Evaluation Metrics. In order to further measure the recommendation performance of different methods, we use different evaluation metrics to evaluate them. As illustrated in Figure 6, the results show that as the value of K increases, the recommendation accuracy rate and recommendation performance decrease, which indicates that the ideal recommendation result ranking position is not high. On the other hand, from the results, we can see that the recommendation performance

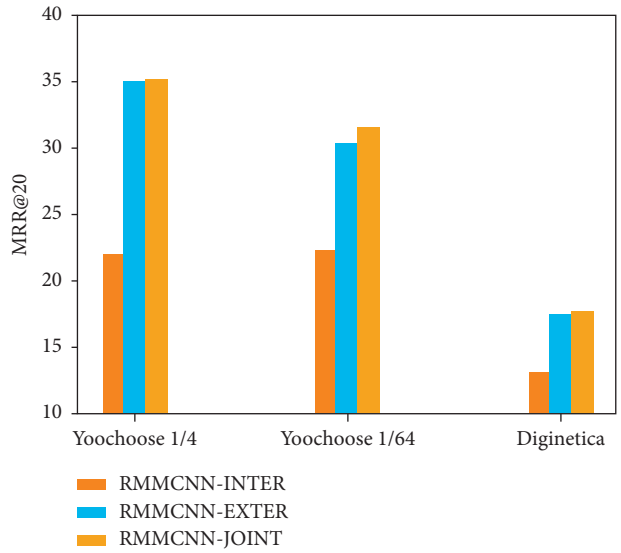


Figure 5: MRR@20 of different connection schemes.

Table 4: The performance of diverse session embeddings (K = 20).

Methods	Internal (%)		External (%)		Joint (%)	
Measures	P@K	MRR@K	P@K	MRR@K	P@K	MRR@K
Yoochoose 1/4	57.20	21.96	67.64	34.90	72.93	35.04
Yoochoose 1/64	58.10	22.22	67.87	30.26	70.94	31.46
Diginetica	37.56	12.97	49.71	17.38	51.82	18.19

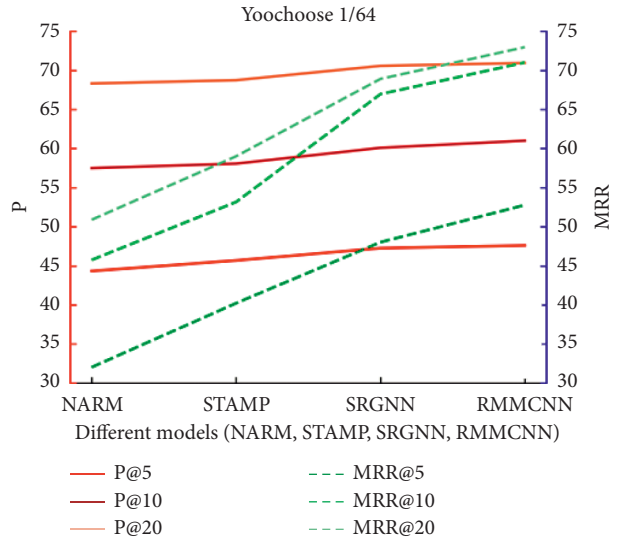


Figure 6: The Precision and MRR of different K settings on Yoochoose 1/64.

improves as the model complexity increases. This shows that as the complexity of the model increases, the embedded information extracted by the model is richer, and the recommendation results rank higher.

We tested on the Yoochoose dataset and Diginetica dataset. On the Yoochoose dataset, the result of P@10 is 60.32% and the result of P@20 is 70.94%. The results are better than those in other baseline models. On the Diginetica dataset, the result of P@10 is 36.64%, which is lower than 39.89% of NARM. However, the result of P@20 is 51.82%, while the result of SRGNN is 50.73%, indicating that the ranking of RMMCNN recommendation results in Diginetica data is mostly in [10, 20].

In the indicators of P@10 in Table 5, we see that NARM is superior to other methods. This is mainly due to the NARM's hybrid encoder attention mechanism. In the

Table 5: The performance of various K (P5 means P@K with $K = 5$, M5 means MRR@K with $K = 5$, etc.).

Datasets	Yoochoose 1/64						Diginetica					
Measures	P5	M5	P10	M10	P20	M20	P5	M5	P10	M10	P20	M20
NARM	44.34	26.21	57.50	27.97	68.32	28.63	27.95	15.67	**39.89**	17.26	49.70	16.17
STAMP	45.69	27.26	58.07	28.92	68.74	29.67	20.41	10.87	30.46	12.20	45.64	14.32
SR-GNN	47.26	28.26	60.09	30.69	70.57	30.94	26.64	15.10	37.91	16.61	50.73	17.59
RMMCNN	**47.61**	**28.87**	**60.32**	**30.71**	**70.94**	**31.46**	**26.83**	**15.90**	36.64	**17.93**	**51.82**	**18.19**

Diginetica dataset, the attention mechanism is more complex than those in other models. More representation information is aggregated, so the performance is significantly better than those of other recommendation models. However, as the K value increases, the recommendation performance of NARM is not as good as those of other more complex models. We argue that more complex models introduce more dimensional representation information, which is much better than simply using the attention mechanism.

5. Conclusions

In this paper, we first propose a novel multichannel convolution model to capture the rich information of items for recommendation. Then, we use the attention mechanism to obtain the features of the user clicked items sequence adaptively and combine the internal and external features of the session to jointly generate the final users' session vector embedding. Once we have the vector, we perform a linear transformation along with the softmax function. We convert the result into a probability value between $[0, 1]$. We have conducted some experimental results with stat-of-the-art recommender systems with two real datasets and both Precision and MRR have been improved. In the future, we will further study the richer representations of items' embedding to make more accurate recommendations.

Data Availability

The data used can be found at https://2015.recsyschallenge.com/challenge.html and https://competitions.codalab.org/competitions/11161#learn_the_details-data2.

Conflicts of Interest

There are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Major Project of the National Natural Science Foundation of China (no. 51935002) and the National Key Research and Development Program of China (no. 2018YFC0831903).

References

[1] Z. Y. Ji, H. Y. Pi, and W. Yao, "A hybrid recommendation model based on fusion of multi-source heterogeneous data," *Journal of Beijing University of Posts and Telecommunications*, vol. 42, no. 1, pp. 126–132, 2019.

[2] D. R. Liu, Y. C. Chou, and C. T. Jian, "Integrating collaborative topic modeling and diversity for movie recommendations during news browsing," *Kybernetes*, vol. 49, no. 11, pp. 2633–2649, 2019.

[3] Y. Hu, F. Xiong, D. Lu, X. Wang, X. Xiong, and H. Chen, "Movie collaborative filtering with multiplex implicit feedbacks," *Neurocomputing*, vol. 398, pp. 485–494, 2020.

[4] H.-Y. Chang, S.-C. Huang, and J.-H. Wu, "A personalized music recommendation system based on electroencephalography feedback," *Multimedia Tools and Applications*, vol. 76, no. 19, pp. 19523–19542, 2017.

[5] J. W. Chang, C. Y. Chiou, J. Y. Liao et al., "Music recommender using deep embedding-based features and behavior-based reinforcement learning," *Multimedia Tools and Applications*, Springer, Berlin, Germany, 2019.

[6] F. M. Belém, R. M. Silva, C. M. V. de Andrade et al., "Fixing the curse of the bad product descriptions"—search-boosted tag recommendation for E-commerce products," *Information Processing & Management*, vol. 57, no. 5, Article ID 102289, 2020.

[7] K. Wang, T. Zhang, T. Xue, Y. Lu, and S.-G. Na, "E-commerce personalized recommendation analysis by deeply-learned clustering," *Journal of Visual Communication and Image Representation*, vol. 71, Article ID 102735, 2020.

[8] M. Ludewig and D. Jannach, "Evaluation of session-based recommendation algorithms," *User Modeling and User-Adapted Interaction*, vol. 28, pp. 331–390, 2018.

[9] L. M. Zhang, P. Liu, and J. A. Gulla, "Dynamic attention-integrated neural network for session-based news recommendation," *Machine Learning*, vol. 108, pp. 1851–1875, 2019.

[10] M. Shi, J. X. Liu, and D. Zhou, "A hybrid approach for automatic mashup tag recommendation," *Journal of Web Engineering*, vol. 16, pp. 676–692, 2017.

[11] V. Maihami, D. Zandi, and K. Naderi, "Proposing a novel method for improving the performance of collaborative filtering systems regarding the priority of similar users," *Physica A: Statistical Mechanics and Its Applications*, vol. 536, Article ID 121021, 2019.

[12] C. Zou and Z. Chen, "Joint latent factors and attributes to discover interpretable preferences in recommendation," *Information Sciences*, vol. 505, pp. 498–512, 2019.

[13] J. Liu, L. Fu, X. Wang, F. Tang, and G. Chen, "Joint recommendations in multilayer mobile social networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 10, pp. 2358–2373, 2020.

[14] J. H. Wang and T. W. Liu, "Improving sentiment rating of movie review comments for recommendation," in *Proceedings of the IEEE International Conference on Consumer Electronics—Taiwan (ICCE-TW)*, pp. 12–14, Taipei, Taiwan, June 2017.

[15] Z. Y. Ji, X. J. Song, H. Y. Pi, and C. Yang, "Recommended model for fusing multi-source heterogeneous data based on

deep learning," *Journal of Beijing University of Posts and Telecom*, vol. 42, no. 6, pp. 35–42, 2019.

[16] W. Zhang, Z. Zhang, S. Zeadally et al., "MASM: a multiple-algorithm service model for energy-delay optimization in edge artificial intelligence," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4216–4224, 2019.

[17] W. Zhang, Z. Zhang, H. C. Chao et al., "Toward intelligent network optimization in wireless networking: an auto-learning framework," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 76–82, 2019.

[18] J. Li, H. Xu, X. W. He et al., "Tweet modeling with LSTM recurrent neural networks for hashtag recommendation," in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 24–29, Vancouver, Canada, July 2016.

[19] J. F. Dong, X. R. Li, C. X. Xu et al., "Feature re-learning with data augmentation for content-based video recommendation," in *Proceedings of the 26th ACM Multimedia Conference (MM)*, pp. 22–26, Seoul, South Korea, October 2018.

[20] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "STAMP: short-term attention/memory priority model for session-based recommendation," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 19–23, London, UK, August 2018.

[21] S. Wu, Y. Y. Tang, Y. Q. Zhu et al., "Session-based recommendation with graph neural networks," 2019, https://arxiv.org/abs/1811.00855.

[22] M. Jiang, Z. Y. Yang, and C. Zhao, "What to play next? A RNN-Based Music Recommendation System," in *Proceedings of the 51st Asilomar Conference on Signals, Systems, and Computers*, pp. 356–358, Pacific Grove, CA, USA, November 2017.

[23] B. Xia, Y. Li, Q. M. Li et al., "Attention-based recurrent neural network for location recommendation," in *Proceedings of the 12th International Conference on Intelligent Systems and Knowledge Engineering (IEEE ISKE)*, Nanjing, CHINA, November 2017.

[24] Y. Cai, S. B. Dong, and J. L. Hu, "Jointly modeling user and item reviews by CNN for multi-domain recommendation," in *Proceedings of the 24th China conference on information retrieval (CCIR)*, September 2018.

[25] R. Gao, H. F. Xia, J. Li et al., "DRCGR: Deep reinforcement learning framework incorporating CNN and GAN-based for interactive recommendation," in *Proceedings of the 19th IEEE International Conference on Data Mining (ICDM)*, Beijing, China, November 2019.

[26] W. Q. Fan, Y. Ma, Q. Li et al., "Graph neural networks for social recommendation," in *Proceedings of the World Wide Web Conference (WWW)*, San Francisco, CA, USA, May 2019.

[27] X. Xian, L. Fang, and S. Sun, "ReGNN: a repeat aware graph neural network for session-based recommendations," *IEEE Access*, vol. 8, pp. 98518–98525, 2020.

[28] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the ACM World Wide Web Conference*, pp. 285–295, April 2001.

[29] W. Cheng, G. Yin, Y. Dong, H. Dong, and W. Zhang, "Collaborative filtering recommendation on users' interest sequences," *PLoS One*, vol. 5, Article ID e0155739, 2016.

[30] W. T. H. Putri, M. S. Prastio, R. Hendrowati et al., "Content-based filtering model for recommendation of Indonesian legal article study case of klinik hukumonline," in *Proceedings of the 4th International workshop on big data and information security (IWBIS)*, Bali, Indonesia, October 2019.

[31] H. Zhang, Y. Sun, M. Zhao, T. W. S. Chow, and Q. M. J. Wu, "Bridging user interest to item content for recommender systems: an optimization model," *IEEE Transactions on Cybernetics*, vol. 50, no. 10, pp. 4268–4280, 2020.

[32] T. Trinh, D. M. Wu, R. L. Wang et al., "An effective content-based event recommendation model," *Multimedia Tools and Applications*, Springer, Berlin, Germany, 2020.

[33] E. Rojsattarat and N. Soonthornphisaj, "Hybrid recommendation: combining content-based prediction and collaborative filtering," in *Proceedings of the 4th International Conference on Intelligent Data Engineering and Automated Learning*, Hong Kong, China, March 2003.

[34] M. Kiewra, "RankFeed—recommendation as searching without queries: new hybrid method of recommendation," *Journal of Universal Computer Science*, vol. 11, no. 2, pp. 229–249, 2005.

[35] M. Kolahkaj, A. Harounabadi, A. Nikravanshalmani et al., "A hybrid context-aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining," *Electronic Commerce Research and Applications*, vol. 42, 2020.

[36] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, Banff, Canada, April 2014.

[37] J. Li, P. J. Ren, Z. M. Chen et al., "Neural attentive session-based recommendation," in *Proceedings of the ACM Conference on Information and Knowledge Management (CIKM)*, Singapore, Singapore, November 2017.

Hindawi
Complexity
Volume 2021, Article ID 8841822, 15 pages
https://doi.org/10.1155/2021/8841822

Research Article

Graph-Based Analysis of RNA Secondary Structure Similarity Comparison

Lina Yang,[1] Yang Liu,[1] Xiaochun Hu,[2] Patrick Wang,[3] Xichun Li,[4] and Jun Wu[5]

[1]*Computer and Electronic Information, Guangxi University, Nanning 530004, China*
[2]*School of Information and Statistics, Guangxi University of Finance and Economics, Nanning 530007, China*
[3]*Computer and Information Science, Northeastern University, Boston 02115, USA*
[4]*Guangxi Normal University for Nationalities, Chongzuo, China*
[5]*School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Yang Liu; 1813302002@st.gxu.edu.cn

Received 14 September 2020; Revised 16 February 2021; Accepted 10 March 2021; Published 23 March 2021

Academic Editor: Jia Wu

In organisms, ribonucleic acid (RNA) plays an essential role. Its function is being discovered more and more. Due to the conserved nature of RNA sequences, its function mainly depends on the RNA secondary structure. The discovery of an approximate relationship between two RNA secondary structures helps to understand their functional relationship better. It is an important and urgent task to explore structural similarities from the graphical representation of RNA secondary structures. In this paper, a novel graphical analysis method based on the triple vector curve representation of RNA secondary structures is proposed. A combinational method involving a discrete wavelet transform (DWT) and fractal dimension with sliding window is introduced to analyze and compare the graphs derived from feature extraction; after that, the distance matrix is generated. Then, the distance matrix is analyzed by clustering and visualized as a clustering tree. RNA virus and noncoding RNA datasets are applied to perform experiments and analyze the clustering tree. The results show that the proposed method yields more accurate results in the comparison of RNA secondary structures.

1. Introduction

The secondary structure of RNA is a double-stranded structure constructed according to the principle of complementary base pairing. In RNA, U (uracil) is complementary to A (adenine) and *G* (guanine) is complementary to C (cytosine). RNA is usually transcribed from DNA and acts as a bridge between DNA and proteins [1]. The versatility of an RNA molecule depends on its secondary structure. RNAs with similar structures tend to have similar functions or properties, but the opposite does not necessarily hold true. Many RNA molecules are conserved at the structural level, but they have little sequence similarity. Therefore, the comparison of RNA secondary structures is key to elucidating their functional and evolutionary relationships.

Most recent studies have focused on RNA secondary structure prediction [2–4], and comparisons of RNA secondary structures have not yet been sufficiently studied. At the present stage, the comparison methods for RNA secondary structures are mainly divided into two types: alignment-based methods and alignment-free methods.

Alignment-based methods mainly rely on an RNA secondary structure represented by a string or tree [5–9].

The Sankoff algorithm used a free-energy minimization method to synchronize the folding and alignment of two or more RNA sequences [9]. Combining the information from folding and alignment optimizes the objective function when processing the problem separately, while generalizing the alignment method to the original sequence reconstruction problem and selecting appropriate, existing constraints for possible solutions. However, this suffers from a problem. The time cost is high and about O (n^3k^N) (N is the number of sequences, n is the maximum sequence length, and the value of k is small enough). In practice, it is

impossible to solve and satisfy both the time complexity problem and the larger memory requirements, which, together with the impractical constraints, are all difficult to achieve.

Therefore, to be more effective in RNA secondary structure alignment, a number of improved algorithms were subsequently developed, such as Consan [10] and, Dynalign [11], PMcomp [12], Stemloc [13], Foldalign [14], locARNA [15], SPARSE [16], MARNA [17], FoldAlignM [18], Murlet [19], CARNA [20], and RAF [21].

Consan is based on Sankoff algorithm and was proposed by Dowell to address two problems not solved by Sankoff: the fact that it cannot be generalized to RNA structures of different lengths and the need to deal with the complexity of the algorithm. First, the method of modelling secondary structures and alignments using random context-independent grammars (SCFG) [22] is extended to construct pair-SCFG descriptions and build unconstrained algorithms. Second, the complete pair-SCFG algorithm is almost infeasible due to the $O(N^4)$ memory requirement and $O(N^6)$ time consumption. The structure algorithm is constrained by adding pins. Long sequence comparison is achieved using anchors fixed partial comparison. Also, the time and space complexity of the algorithm is effectively reduced. However, it is clear that Consan syntax lacks the handling of more complex statistical functions [10].

PMcomp used McCaskill's algorithm to obtain the base-pairing probability matrix and, next, the alignment probability matrix. This algorithm computes the input pairing matrix independently, unlike other combinatorial alignment and folding procedures (e.g., Dynalign and FoldAlign). Promising candidate sequences are preselected using a pattern search procedure, and an algebraic dynamic programming approach is used to scan the candidate sequences. It is folded by McCakill's algorithm, and finally, each individual is compared with each other separately using familiar techniques. The method reached a lightweight energy computation and can be described as simplified Sankoff's algorithm [12].

LocARNA is a novel RNA local comparison tool that has achieved high sensitivity in multiple experiments based on different datasets, but its complexity is still high [15]. For this feature of LocARNA, a sparsified prediction and alignment of RNAs based on their structure ensembles (SRARSE) is introduced. It supports the operation of loop structures for the first time and retains the full flexibility of Sankoff's algorithm. In the benchmark test, high accuracy and speed were both maintained [16].

Non-Sankoff algorithms divide the process into two parts: folding and alignment. Among them, some tree-based methods, such as RNAforester [23], RNAdistance [8], RNAStrAt [24], and RNApdist [25], are widely used.

The RNAforester tree comparison algorithm computes the local similarity in RNA secondary structures. The method provides a new representation of RNA secondary structure in forest representation and provides elaborate dynamic programming implementations using dense two-dimensional tables, which can greatly reduce space requirements in practice. RNAdistance calculates the similarity of the RNA secondary structures by measuring the editing distance of the tree. RNApdist compares the RNA secondary structures based on base-pairing probabilities. The Vienna RNA package can now be used to implement both the RNAdistance and RNApdist methods [26, 27]. Considering the structure with pseudoknots, Michela established ASPRAlign and obtained a method for comparing ASPRA distances [28]. A comparison of algebraic RNA trees and structural RNA trees [29] is made with ignoring the primary sequence and considering only the structure of the molecule. The execution time of the ASPRAlign workbench tool to process molecular pairs was compared experimentally. The experiments show that the time complexity of the algorithm is $O(n^2)$. Notably, these comparison algorithms have high time complexity. General alignment-free methods are usually based on the numerical representations of RNA secondary structures, followed by the development of the many graphical representations of RNA secondary structures. The visualization of an RNA secondary structure using a graphical representation is more intuitive, thus providing a new way of thinking about the comparison of RNA secondary structures. Previously, researchers used eight symbols to represent RNA, but this representation was accompanied by a loss of information [30]. A feature sequence may correspond to diverse RNA secondary structures. That is, the feature sequences obtained by such a method are not unique. To reduce the loss of information, based on a 4×4 matrix of RNA sequences proposed by Randic [31], Yao used four horizontal lines and eight symbols to represent the RNA secondary structure [32]. The method avoids the loss of information due to the crossover and overlap of curves, but the information loss caused by the limitations of feature-invariant extraction still exists. Randic proposed a method to visualize the secondary structure of RNA without loss of information [33]. The method is based on [30] labeling bases before and after hydrogen bonds, using 12 symbols to represent bases at different positions and, in turn, using numbers and graphs to represent RNA sequences with structural information. To illustrate the validity of the method, two RNAs that are similar were selected, and visualization of RNA sequences using the method enables visual observation of the sequence differences. In [34], Li proposed a novel graphical representation of RNA secondary structures (TV-Curve) and introduced a wavelet decomposition to compare RNA secondary structures. Based on the feature representation of RNA secondary structure with eight symbols proposed in [30], each character is represented by three vectors, and the vectors are connected sequentially to generate a unique graphical representation of RNA secondary structure. After feature extraction, RNA similarity is estimated using weighted correlations of wavelet domains at different scales. The method integrates the global and local structures. The effect of the program was verified using 100 RNA sequences from the RFAM database and 9 viruses to obtain a similarity metric closer to the actual data. Based on the approach of [34], Li further created a web server for multiscale similarity RNA secondary structure comparison based on the curve representation of triple vectors (RNA-TV-curve), including

RNA visualization, mutation analysis, and multiple RNA structure comparison [35].

Neutral components (NCs) in the genotype-phenotype (GP) maps were found in [36] to be able to influence the biological sequence function. Thus, based on a method that reveals the hierarchical community structure, which reveals the hierarchical community structure solely from the sequence constraints and composition of the genotypes that form a given NC, community detection has been used to analyze RNA secondary structure, and recent works on community detection have been studied in [37]. Some novel classifiers and learning models have also been proposed for multigraph analysis. Multigraph feature-based representation and learning methods for multigraph classification were used to achieve an effective learning model for multigraph classification [38]. Subsequently, to address the complexity of graph data generalization in multigraph classification, Wu proposed a boosting-based multigraph classification framework (bMGC). The framework uses a two-level weighting strategy to balance label ambiguity and introduces dynamic subgraph selection criteria to solve the structured data representation problem [39].

Besides, the classification and comparison of noncoding RNAs is also gaining attention. The deep-learning-based model ncRDeep is a recently proposed tool to efficiently classify ncRNA families [40]. It uses a simple convolutional neural network and RNA sequence information only to predict the class of ncRNAs. The experiments in the article evaluate the performance of ncRDeep using a benchmark dataset. Ultimately, the method was effective in improving the accuracy of prediction.

Wavelet analysis is a synthesis of ideas in pure mathematics, physics, and engineering. A commonly used windowed Fourier transform, such as the short-time Fourier transform (STFT), analyzes the signal with a fixed sliding window. Obviously, fixed-length sliding window processing is not suitable for all signals. The linear time-frequency analysis for nonstationary signals should have different resolutions at different positions in the time-frequency plain; in other words, it should be a multiresolution analysis method. Wavelet transform is a multiresolution analysis method. Furthermore, wavelet is a fairly simple and effective mathematical tool that has been widely used. A combination of wavelet transform and neural network has been proposed to achieve highly accurate machine fault diagnosis [41]. In [42], wavelet analysis was used to identify membrane protein types. It was also used to realize RNA secondary structure similarity analysis. In practical applications, especially when implementing the wavelet transform on a computer, the signal must be discretized and analyzed by the discrete wavelet transform. By detecting and analyzing protein secondary structures by discrete wavelet transforms, the correlation of the amino acid sequence and secondary structure was tested by the hydrophobic value of the amino acids [43].

The fractal dimension (FD) reflects the validity of the space occupied by a complex form, which is a measure of the irregularity of a complex form [44]. FD acknowledges that various parts of the world may be similar in some way to the whole region under certain conditions or processes, and it recognizes that changes in spatial dimensions can be discrete and continuous [45]. Reference [46] combines empirical modal decomposition and multifractal detrended fluctuation analysis to study the fractal characteristics of harmonic signals. Additionally, the fractal dimension has been introduced in the studies of biological molecules. Yu exploited the hydrophobicity of amino acids and fractal analysis to classify the protein structure [47]. In [48], Yang applied the fractal dimension to protein sequence similarity analysis and obtained a high degree of similarity. Yang also integrated the fractal dimension with empirical mode decomposition to compare protein sequences and optimise the feature extraction process. Experiments show that the combination of the two methods extracts higher quality features [49]. The fractal dimension has also been recently used to determine the distribution of purines and pyrimidines in the miRNAs of humans, gorillas, chimpanzees, mice, and rats [50].

In this paper, we propose a novel combined algorithm. Based on the characteristics of discrete wavelet transform (DWT) multiresolution analysis and the universality of FD, DWT is used to analyze the graphical representation of RNA secondary structures and FD is used to quantitatively characterize their geometric structure features. The purpose of this paper is to explore the application of DWT and fractal dimension in the comparison of RNA secondary structures.

The paper is organized as follows. We outline the theory of the fractal dimensions and the graphical characterization of RNA secondary structure in Section 2. Numerical experiments and the results are then given and discussed in Section 3. Section 4 is the conclusion.

2. Materials and Methods

In this paper, a combination of the discrete wavelet transform, detrended fluctuation analysis, and sliding window is proposed to evaluate the similarity of RNA secondary structures based on the TV-curve of RNA. The TV-curve graphical representation of RNA is used to construct the feature vector, which is analyzed by the discrete wavelet transform, and then, the sliding window is introduced. For a signal of fixed length within the window, the fractal dimension is calculated using the detrended fluctuation analysis (DFA) method to obtain the distance matrix. Finally, the clustering analysis is performed, and the corresponding cluster analysis graph is obtained. The algorithmic process is shown in Figure 1.

2.1. Basic Concepts of Fractal Dimension

2.1.1. Theoretical Fractal Dimension. The German mathematician F. Hausdorff studied the properties and quantities of singular aggregates and proposed the concept of fractal dimension called the Hausdorff dimension in *1918* [51]. It is one of the most important fractal dimensions. The Hausdorff dimension can be used for any set. The theory of the Hausdorff dimension is as follows:

Suppose a nonempty subset A of an n-dimensional Euclidean space $\mathbb{R}^n$, where the diameter of A is the maximum distance of any two points in A; then,

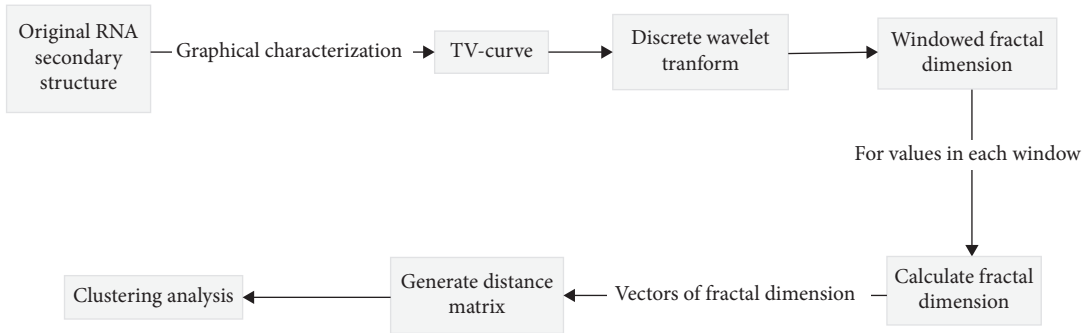


Figure 1: The flowchart of our method.

$$|A| = \sup\{|x - y|: \ x, y \in A\}, \tag{1}$$

where sup $\{\cdot\}$ indicates the supremum of $\{\cdot\}$. If $C \subset \cup_{i=1}^{\infty} A_i$ and $0 \le |A_i| \le \delta$, for any i, $\{A_i\}$ is called a δ-cover of C.

Assuming $C \subset \mathbb{R}^n$ and $0 \le s \le \infty$, for any $\delta > 0$,

$$H_\delta^s(C) = \inf\left\{\sum_{i=1}^{\infty} |U_i|^s: \ \{A_i\} \text{ is a } \delta - \text{cover of } C\right\}. \tag{2}$$

The abovementioned equation refers to the coverage of C whose diameter does not exceed δ, and the sum of these diameters is minimized to the power of s. The symbol $\inf\{\cdot\}$ takes the minimum value of $\{\cdot\}$.

As δ decreases, the clusters covering C in equation (2) also decreases, and the infimum $H_\delta^s(C)$ increases accordingly. Moreover, when $\delta \longrightarrow 0$, it tends to a limit, denoted as

$$H^s(C) = \lim_{\delta \longrightarrow 0} H_\delta^s(C). \tag{3}$$

For any subset C in $\mathbb{R}^n$, the limit exists, and the limiting value is commonly 0 or ∞, and $H^s(C)$ is called the s-dimensional *Hausdorff measure* of C.

In equation (2), for a given set $C \subset \mathbb{R}^n$ and $\delta < 1$, $H_\delta^s(C)$ does not increase with respect to s, so it can be shown from equation (3) that $H^s(C)$ does not increase either. Furthermore, if $t > s$ and $\{A_i\}$ is the δ-cover of C, then

$$\sum_i |A_i|^t \le \sum_i |A_i|^{t-s} |A_i|^s \le \delta^{t-s} \sum_i |A_i|^s. \tag{4}$$

Taking the infimum,

$$H_\delta^t(C) \le \delta^{t-s} H_\delta^s(C). \tag{5}$$

In the case of $t > s$, let $\delta \longrightarrow 0$; if $H^s(C) < \infty$, then $H^t(C) = 0$; thus, the existence of a critical point of s makes $H^s(C)$ "jump" from ∞ to 0. This critical value is known as the Hausdorff dimension of F, recorded as $\dim_H C$, and occasionally called the *Haus do r f f − Basicovitch* dimension.

The definition is expressed as

$$\begin{aligned} \dim_H C &= \inf\{s \ge 0: \ H^s(C) = 0\}, \\ &= \sup\{s: \ H^s(C) = \infty\}. \end{aligned} \tag{6}$$

Therefore,

$$H^s(C) = \begin{cases} \infty, & \text{if } 0 \le s < \dim_H C, \\ 0, & \text{if } s > \dim_H C. \end{cases} \tag{7}$$

In addition, if $s = \dim_H C$, $H^s(C)$ satisfies

$$0 \le H^s(C) \le \infty. \tag{8}$$

2.2. Algorithms of Fractal Dimension Calculation. Although the fractal dimension is widely defined, it is subject to some limitations in practical applications. Due to its high computational complexity and the discrete and finite nature of the scale factor, we usually use a number of algorithms to approximate the fractal dimension. To date, many methods have been developed to compute fractal dimensions, such as Katz's algorithm [52], Petrosian's algorithm [53], Higuchi's algorithm [54], multifractal detrended fluctuation analysis (MFDFA) [55], detrended fluctuation analysis (DFA) [56], and fluctuation analysis (FA) [57]. In terms of processing time, Katz's algorithm and the DFA method are relatively slow. However, in terms of processing accuracy, Higuchi's algorithm, detrended fluctuation analysis (DFA), and fluctuation analysis (FA) are relatively accurate. In the following, we will introduce Higuchi's algorithm, fluctuation analysis (FA), and detrended fluctuation analysis (DFA).

2.2.1. Higuchi's Algorithm. Treating an RNA sequence as a one-dimensional signal $s_1, s_2, \ldots, s_N$, we construct a subsequence as

$$s_m^k = \left\{s(m), s(m+k), s(m+2k), \ldots, s\left(m + \left\lfloor \frac{N-m}{k} \right\rfloor k\right)\right\}, \quad m = 1, \ldots, k, \tag{9}$$

where m indicates the starting position of the signal and k is the measurement scale. $\lfloor (N-m)/k \rfloor$ means the integer part of $\lfloor (N-m)/k \rfloor$, which is the number of terms in s_m^k.

We calculate the average length $L_m(k)$ of the curve with starting position m according to the measurement scale k.

For $m = 1, \ldots, k$,

$$L_m(k) = \frac{\sum_{i=1}^{\lfloor (N-m)/k \rfloor} |s(m+ik) - s(m+(i-1)k)|\,(N-1)}{\lfloor (N-m/k) \rfloor}, \tag{10}$$

where N is the length of the signal and $(N-1)/(((N-m)/k)\,k)$ is the normalization term. With the given measurement scale, the approximate length of the signal with starting position m is computed, $m = 1, \ldots, k$.

Thus, the approximated signal length can be obtained by

$$L(k) = \frac{1}{k} \sum_{m=1}^{k} L_m(k), \quad k = 1, \ldots, k_{\max}. \tag{11}$$

Finally, the fractal dimension h^* of the signal is calculated by the least-squares method [58]:

$$h^* = \arg\min_h \sum_{k=1}^{K} \left(h * \log\left(\frac{1}{k}\right) - \log(L(k)) + c \right)^2, \tag{12}$$

where c is the bias.

2.2.2. Fluctuation Analysis (FA) Method. The fluctuation analysis (FA) method is commonly used to calculate the Hurst parameter for time series. It works as follows.

Suppose $X = \{X_i, i = 1, 2, \ldots, N\}$ is a random process with a mean value of μ and a variance of σ^2. First, we remove the mean from the signal and represent the new signal as $x = \{x_i, i = 1, 2, \ldots, N\}$, where $x_i = X_i - \mu$.

Then, we construct a new one-dimensional signal $y = \{y_n, n = 1, 2, \ldots, N\}$ with the sum of the first n terms of x,

$$y(n) = \sum_{i=1}^{n} x_i. \tag{13}$$

We test whether the following formula satisfies the power-law formula:

$$F^{(2)}(m) = \langle |y(n+m) - y(n)|^2 \rangle^{1/2} \sim m^H. \tag{14}$$

That is, we test whether $\log_2(F^{(2)}(m))$ and $\log_2(m)$ are satisfied:

$$\log_2\left(F^{(2)}(m)\right) = H * \log_2(m) + c, \tag{15}$$

where c is a constant.

From the abovementioned equation, we can obtain the Hurst parameter H.

In the end, we can obtain FD by the following formula:

$$\mathrm{FD} = 2 - H. \tag{16}$$

2.2.3. Detrended Fluctuation Analysis (DFA) Method. Owing to the complicated characterization of long-range correlations and power-law statistics that RNA TV-curves have, it follows that comprehensively describing and studying the internal features of nonstationary signals such as RNA TV-curves by traditional methods is difficult. For complex, highly nonstationary signals, time series analysis methods derived from statistical physics are currently used. The DFA method can effectively eliminate various unknown trends in the signal, thus avoiding the interference caused by noise and signal and instability.

The DFA method has also been widely used in image processing and other fields [59]. The DFA method has been proven to be an effective method to analyze nonstationary signals. In summary, as the FA, the signal x_r, $r = 1, \ldots, N$ is preliminarily processed, and the mean value x is removed. The summation of i terms before the series x_r as the term i in the new sequence is

$$Y(i) = \sum_{r=1}^{i} [x_r - < x >], \quad i = 1, 2, \ldots, N. \tag{17}$$

Next, the signal *Y(i)* is equally split into data boxes of length k. k, which is the measurement scale, is the number of points per data box.

In each box of length k, a least-squares line is fit to the data. As a result, fitting data box $y_k(j)$ is obtained. Then, we detrend the signal Y.

$$\widetilde{Y_k(j)} = Y(j) - y_k(j). \tag{18}$$

In the end, the different data obtained for each box are integrated and averaged.

$$F_2(k) = \left[\frac{1}{N} \sum_{j=1}^{N} F_{\mathrm{DFA}}^2(k) \right]^{1/2}, \tag{19}$$

where $F_{\mathrm{DFA}}^2(k) = (1/k) \sum_{j=1}^{k} \widetilde{Y_k}^2(j)$.

The scaling exponent α is the slope of the line of fit of $F_2(k)$ and k.

The fractal dimension is calculated by

$$\mathrm{FD} = 2 - \alpha. \tag{20}$$

2.3. Graphical Characterization of RNA Secondary Structure

2.3.1. Dot-Bracket Representation. The sequence of RNA consists of the bases *C*, *G*, *U*, and *A*. The primary sequence does not contain structural information. During the exploration of the RNA function, two bases combine with each other to form base pairs. Typically, the base pairs are *A*-*U*, *G*-*C*. Thus, the secondary structure of RNA is formed. Currently, "*the dot-bracket*" representation is commonly used to signify the secondary structure of RNA, and the dissociative bases and base pairs are indicated by dots and parentheses, respectively. Among them, open brackets express the base pairs near the 50-terminal of the RNA chain, while closed brackets indicate the bases close to the 30-terminal. This representation can be obtained via the Vienna RNA package. However, this linear representation is degenerate in some cases in which different RNA secondary structures have the same characteristic sequence. Nevertheless, this

circumstance usually occurs in short RNA sequences, which can be ignored in the analysis of long and complex RNA secondary structures.

2.3.2. The TV-Curve Representation of RNA Secondary Structure. Based on the "dot-bracket" representation introduced previously, in this part, we will show a method for the feature extraction of RNA secondary structures, called "RNA triple vector curve" representation. This method was proposed by Li in 2012, as a 2-D graphical representation of RNA secondary structures, in which RNA sequence information and structural information are considered [34]. The bases in the RNA secondary structure are divided into two types: nucleotide bases paired by hydrogen bonds and unpaired nucleotide bases. The four unpaired nucleotide bases are denoted by C (cytosine), G (guanine), U (uracil), and A (adenine); in addition, C', G', U', and A' denote paired nucleotide bases. As shown in Figure 2, each of the eight symbols is described by three vectors.

$$\begin{aligned}
&(1,-1),(1,1),(1,-1)\Rightarrow C,\ (1,1),(1,-1),(1,-1)\Rightarrow C',\\
&(1,-1),(1,-1),(1,-1)\Rightarrow G,\ (1,1),(1,1),(1,-1)\Rightarrow G',\\
&(1,1),(1,-1),(1,1)\Rightarrow U,\ (1,-1),(1,1),(1,1)\Rightarrow U',\\
&(1,1),(1,1),(1,1)\Rightarrow A,\ (1,-1),(1,-1),(1,1)\Rightarrow A'.
\end{aligned} \tag{21}$$

The feature sequence is read from 50-terminal to 30-terminal, and the vectors are sequentially connected to obtain the TV-curves. For example, Figure 3 demonstrates the secondary structure of tRNA_Y00055.1 : 4327–4494 and the corresponding TV-curve. An RNA sequence of length N generates a TV-curve with an X-axis length of 3N. An RNA sequence and its secondary structure may produce only one TV-Curve, and similarly, a TV-curve may represent only one RNA secondary structure. Moreover, the TV-curve has no information missing from the process of feature extraction of the RNA.

2.4. Comparison of RNA Secondary Structures Based on the RNA TV-Curve by Detrended Fluctuation Analysis

2.4.1. Discrete Wavelet Transform. The wavelet transform decomposes data, functions, or operators into components of different frequencies and then studies each component on the corresponding scale [60, 61]. We now study the wavelet transform within the range of signal analysis. Similar to the discrete Fourier transform, the discrete wavelet transform is also a time-frequency description method. The signal is decomposed into two characters: the approximation coefficients (AC) and the detailed coefficients (DC).

For the signal $x(t)$, we use two functions at the same time, wavelet function $\Psi(x)$ and scaling function $\Phi(x)$. $\Phi(x)$ for the $s(t)$ overviews approximation, and $\Psi(x)$ details the approximate function of $x(t)$. $\Phi(x)$ can be formulated as

$$\Phi(x) = \sqrt{2}\sum_{n} h_n \Phi(2x-n), \tag{22}$$

where h_n is the low-pass filter. The wavelet function$\Psi(x)$ can be calculated by

$$\Psi(x) = \sqrt{2}\sum_{n} g_n \Phi(2x-n), \tag{23}$$

where gn is the high-pass filter

Two sets of orthogonal functions can be extracted from the wavelet transform: shifted wavelet functions $\Psi(x-k)$ and scaling functions $\Phi(x-k)$.

For any signal $x = \left\{a_k^j\right\}$, the approximation coefficients (AC) and detailed coefficients (DC) in the next level can be quickly calculated as follows:

$$\begin{aligned}
a_k^{j+1} &= \sum_{i\in Z} a_i^j \overline{h}_{i-2k}, \quad j = 0,1,2,\ldots,\\
d_k^{j+1} &= \sum_{i\in Z} a_i^j \overline{g}_{i-2k}, j = 0,1,2,\ldots.
\end{aligned} \tag{24}$$

The decomposition is taken level by level. The lengths of AC and DC obtained after the progressive decomposition are always half the length of the level sequence antecedently. The processing of RNA secondary structures using DFA allows for the description of the overall trends and general characteristics of the signal, as well as the local details.

2.4.2. Windowed Detrended Fluctuation Analysis (DFA) Method. If the fractal dimension is calculated by the DFA of the whole signal, only one scalar can be obtained. The sliding window only calculates the signal values in a fixed-length window [62]. We introduce a sliding window that moves along the signal and calculates the fractal dimension within the window. Calculating the fractal dimensions along a set sliding window yields a feature vector, not just a number. The length of this feature vector is $N-k-1$, where N is the signal length and k is the window width. M indicates the starting position of the signal. A least-squares line is fit to the data in each window of length k. As a result, a fitting data box $y_k^m(j)$ is obtained. Then, we detrend signal Y:

$$\widetilde{Y}_k(j) = Y(j) - y_k^m(j). \tag{25}$$

For each window, we calculate the following formula:

$$F^2_{\mathrm{DFA}_m}(k) = \frac{1}{k}\sum_{j=1}^{k} \widetilde{Y}_k^{\,2}(j). \tag{26}$$

Next, the average is calculated:

$$F_2(k) = \left[\frac{1}{N}\sum_{j=1}^{N} F^2_{\mathrm{DFA}_m}(k)\right]^{1/2}. \tag{27}$$

Transform the value of k, plot the *lb-lb* graph of $F_2(k) - k$, fit the straight line, calculate the slope, and obtain the scaling exponent α. The fractal dimension is calculated by

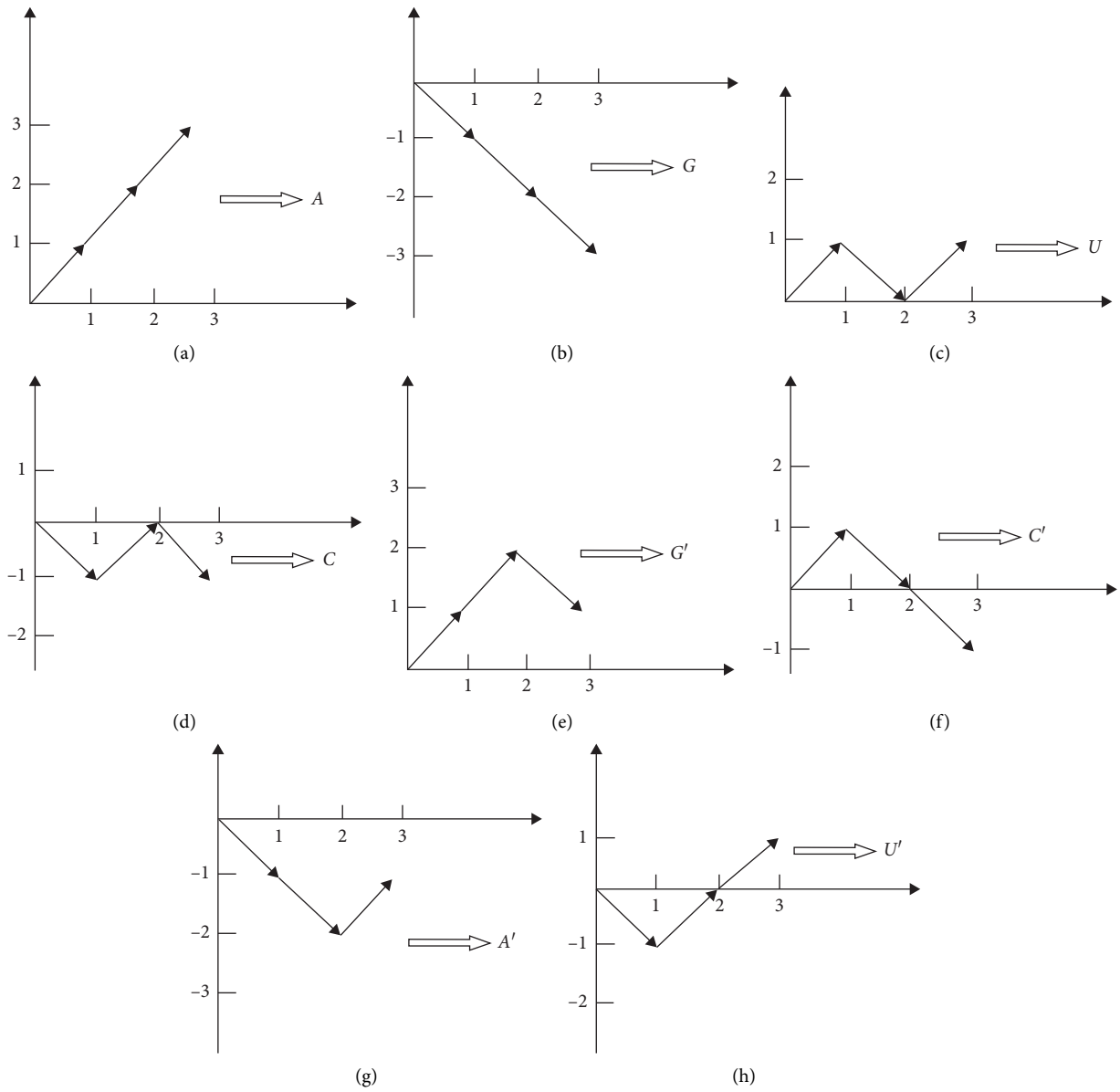


Figure 2: Graphical representation of the eight nucleotides.

$$FD = 2 - \alpha. \tag{28}$$

3. Results and Discussion

3.1. Results. In this section, the optimal experimental setting was obtained by the trial-and-error method: the level of wavelet decomposition was set to 3 and the window width was set to 17. Figure 4 shows the DWT and DFA on the TV-curve graphical representation of a RNA secondary structure using our method. Nine commonly used viruses were selected in this experiment. These viruses have conserved secondary structures, and the differences between them are subtle enough to verify the effectiveness of the experimental method in the classification of secondary structures with slight differences. Information on these nine viruses is presented in Figure 5. The nine viruses include the alfalfa mosaic virus (ALMV), apple mosaic virus (APMV), citrus leaf rugose virus (CiLRV), citrus variegation virus (CVV), elm mottle virus (EMV), lilac ring mottle virus (LRMV), prune dwarf ilarvirus (PDV), asparagus virus II (AVII), and tobacco streak virus (TSV) [63].

The distance matrix of the nine viruses obtained by the method of this paper is shown in Table 1. Next, as shown in Figure 6, the proposed method is compared with the results obtained by MEGA software [64], while the classification effect was compared with classical methods such as RNApdist and the innovative methods ASPRA distance and Li's method. The cluster analysis graph generated by the detrended fluctuation analysis method is based on the distance matrix calculated in MATLAB. In MEGA software, the maximum likelihood method is used to generate the cluster analysis graph.

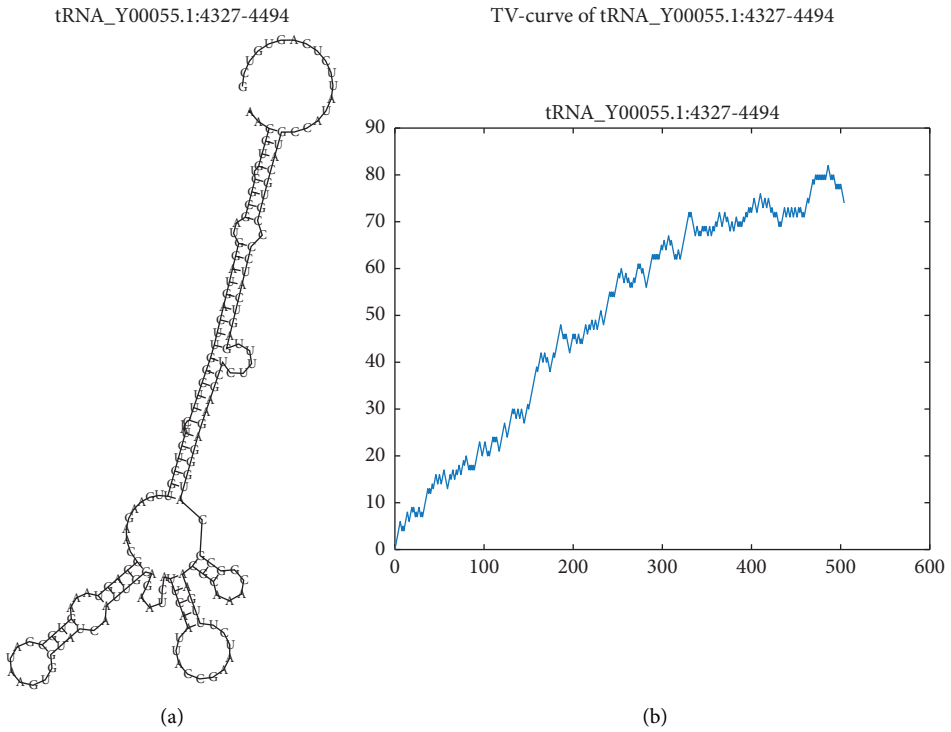


Figure 3: The TV-curve of tRNA_Y00055.1 : 4327–4494.

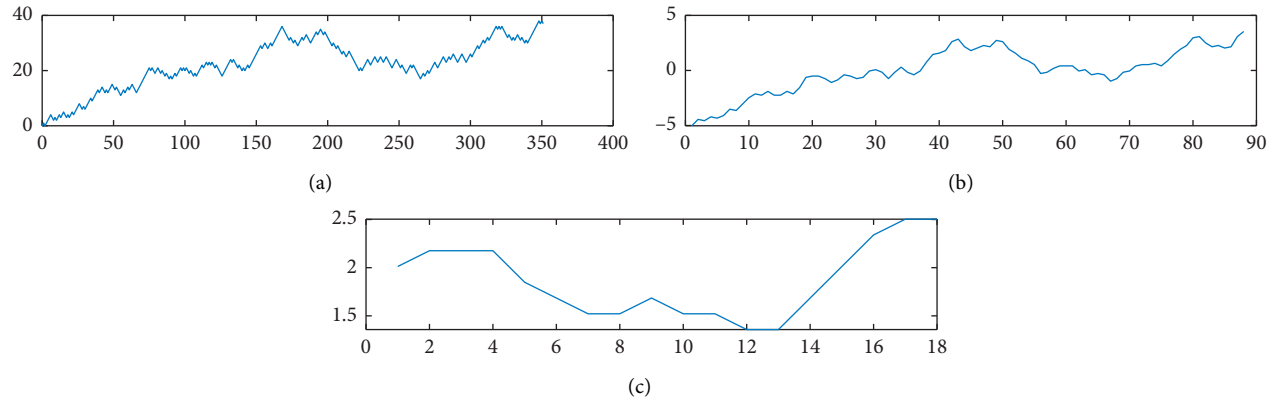


Figure 4: (a) Graphical representation of RNA secondary structure. (b) The third-level discrete wavelet transform. (c) The DFA fractal dimension of (b) obtained with a sliding window width of 17.

In addition, an experiment containing 11 noncoding RNAs is performed to test the applicability of our method in comparing the similarities between noncoding RNAs. Eight sequences are randomly selected from the ncRNAs in the RFAM database; additionally, three human ncRNA sequences from the NONCODE database are randomly sampled for the experiment. The information of these ncRNAs is provided in Table 2, and the distance matrix generated by the experiment is shown in Table 3. Based on this dataset, the results of the classification using ncRDeep are shown in Table 4.

Unfortunately, ncRDeep does not provide the corresponding distance matrix as well as the cluster analysis graph like the other methods, which would not provide a complete comparison in this paper. The remaining experimental results, including the proposed method, RNApdist, ASPRAlign, MEGA software, and Li's method, are shown in Figure 7.

To get rid of the specificity, we randomly extracted 100 ncRNA sequences from four families in the RFAM database, including tRNA, RNase P arch, miRNA, and 5s rRNA. We constructed three cluster analysis graphs using our method, Li's method, RNApdist, and ASPRAlign (see Figure 8: Figure S1, Figure S2, and Figure S3). As shown in Figure 8, the new dataset was applied to build a cluster analysis graph using the proposed method. Four branches can be clearly shown in Figure 8, where the 5 s rRNA family is thoroughly distinguished, while four ncRNAs in tRNA failed to successfully assign with most of the same family. As well, in the

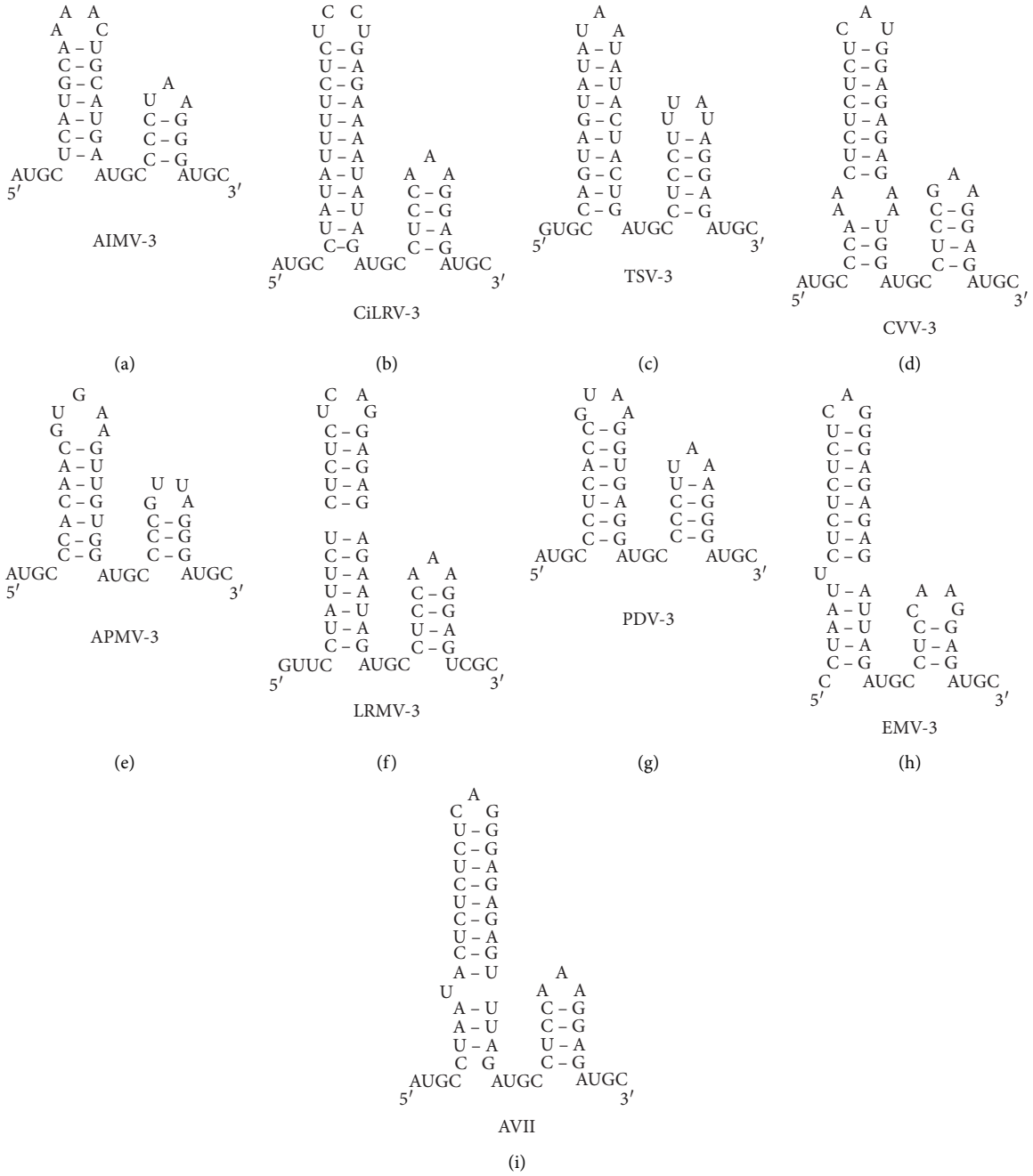


Figure 5: Secondary structure information of nine viruses.

miRNA and RNase P arch families, there are small numbers of ncRNAs that are distant from their major branches. However, among the method proposed by us, RNApdist, Li's method, and ASPRAlign, the classification effect of the proposed method is clearly superior to that of the other three methods.

3.2. Discussion. A detailed explanation of the cluster analysis graphs is given below; it is apparent from Figure 6 that the cluster analysis graph shown in Figure 6(a) is more similar to the standard cluster analysis graph (Figure 6(e)) generated by MEGA software. First, AVII and APMV and ALMV and PDV have a high similarity, which is also clearly reflected in Figure 6(a). However, it is not reflected in the other cluster analysis graphs, such as Figures 6(b)–6(d), and they deviate from the standard cluster analysis graph generated by MEGA software. Obviously, there are three branches in Figure 6(e). CiLRV is closer to LRMV, EMV, and TSV than the remaining four, and Figures 6(a) and 6(b) have similar representations which is not reflected in Figures 6(c) and 6(d). In addition, EMV is closer to LRMV and CVV in all viruses, which is visible at a glance in Figure 6(e), while Figure 6(a) has the same branch. They are, however, separated in Figures 6(b) and 6(c), as well as in Figure 6(d). In

Table 1: The distance matrix of nine viruses generated by our method.

	ALMV	APMV	CiLRV	TSV	CVV	PDV	LRMV	EMV	AVII
ALMV	0	0.718	0.641	0.769	0.667	0.538	0.692	0.846	0.692
APMV		0	0.718	0.718	0.667	0.846	0.718	0.821	0.641
CiLRV			0	0.692	0.462	0.718	0.487	0.718	0.590
TSV				0	0.744	0.795	0.744	0.769	0.769
CVV					0	0.769	0.282	0.615	0.718
PDV						0	0.718	0.744	0.718
LRMV							0	0.590	0.692
EMV								0	0.846
AVII									0

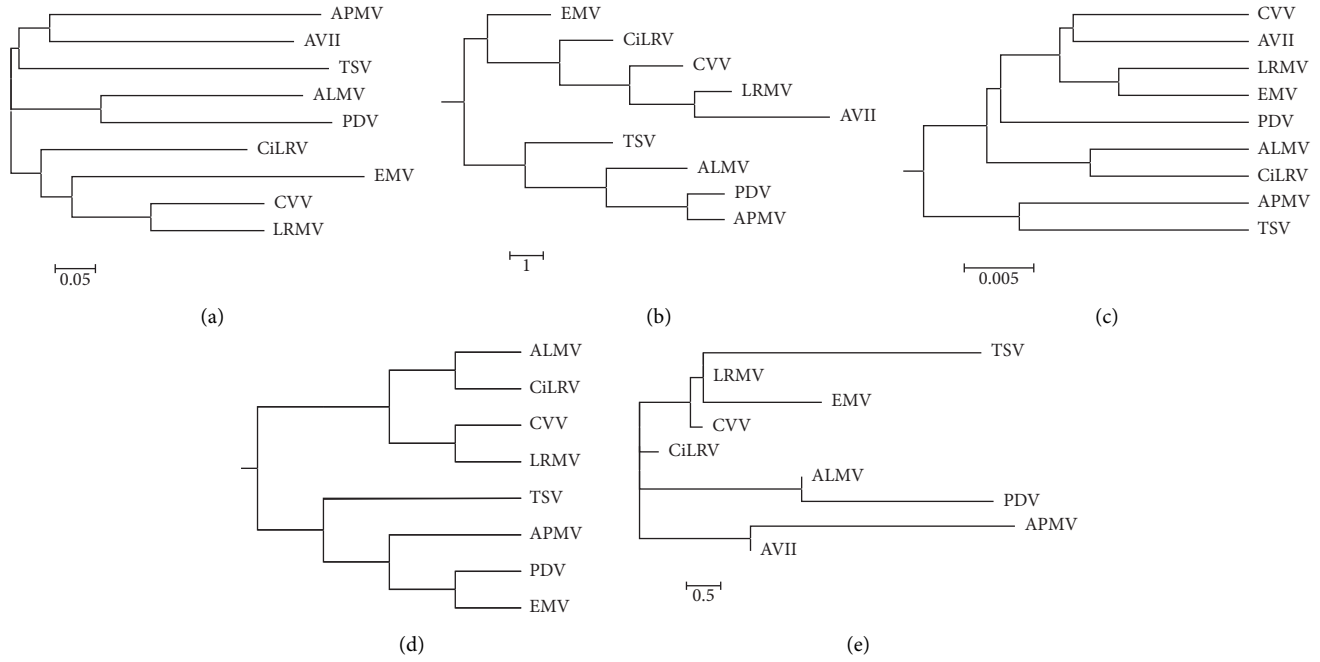


Figure 6: The cluster analysis graph of nine virus sequences generated by (a) our method, (b) Li's method, (c) RNApdist, (d) ASPRAlign, and (e) MEGA software.

Table 2: 11 noncoding RNA sequences from the RFAM and NONCODEv5 databases.

Name	Length	Family
NONHSAT000002.2	1653	NONCODEv5_human
NONHSAT000003.2	1483	
NONHSAT000004.2	632	
MF489813.1	849	tRNA
MF489812.1	535	
MF489811.1	549	
MF489810.1	566	
MF489809.1	549	
MF489808.1	552	
MF489806.1	554	
NT_033777.3	299	RNase P RNA

conclusion, compared with other methods, the proposed method can obtain the similarities of RNA secondary structures more accurately.

The experimental results of the same dataset, the cluster analysis graphs, and the distance between them are measured using the classical Robinson–Foulds (RF) metric [65]. In [65], the smallest number of steps to transform the topology of one of the trees into the other tree by the necessary operation (α or α^{-1}) is noted as the distance between the two trees.

α is the contraction of disjoint, possibly empty set, or zero distance nodes with other normal nodes into a single node while the label changes.

$$T_2 = \alpha(T_1, p_r p_s), \tag{29}$$

where trees T_1 and T_2 have the same set of species, $p_r p_s$ is the edge, and T_2 is obtained from T_1 by shrinking the edge $p_r p_s$.

In turn, α^{-1} is the inverse operation of α, which means dispersing a node into two nodes.

In this paper, the metric was used to calculate the distance between evolutionary trees. Thus, the classification efficiency of the method is verified.

The distance between evolutionary trees generated by the different methods and those generated by MEGA software is

Table 3: The distance matrix of eleven ncRNA sequences produced by our method.

	NONHSAT000002.2	NONHSAT000003.2	NONHSAT000004.2	MF489813.1	MF489812.1	MF489811.1	MF489810.1	MF489809.1	MF489808.1	MF489806.1	NT_033777.3
NONHSAT000002.2	0	0.3882	0.4843	0.5123	0.5178	0.5153	0.5216	0.5212	0.5279	0.5394	0.5510
NONHSAT000003.2		0	0.4571	0.5337	0.5505	0.5504	0.5148	0.5173	0.5160	0.5252	0.5636
NONHSAT000004.2			0	0.5184	0.4199	0.4214	0.4698	0.5221	0.5199	0.4999	0.5348
MF489813.1				0	0.4989	0.4802	0.4905	0.4806	0.4824	0.5334	0.5929
MF489812.1					0	0.0204	0.3603	0.3634	0.4206	0.4557	0.4776
MF489811.1						0	0.3597	0.3519	0.4133	0.4616	0.4736
MF489810.1							0	0.3049	0.4157	0.4632	0.5165
MF489809.1								0	0.3731	0.4445	0.4902
MF489808.1									0	0.4139	0.4933
MF489806.1										0	0.4591
NT_033777.3											0

Table 4: The classification results of ncRDeep.

Sequence ID	Result
NONHSAT000002.2	Ribozymes
NONHSAT000003.2	Ribozymes
NONHSAT000004.2	Ribozymes
MF489813.1	Intron_gpl
MF489812.1	Intron_gpl
MF489811.1	Intron_gpl
MF489810.1	Intron_gpl
MF489809.1	Intron_gpl
MF489808.1	Intron_gpl
MF489806.1	Intron_gpl
NT_033777.3	Ribozymes

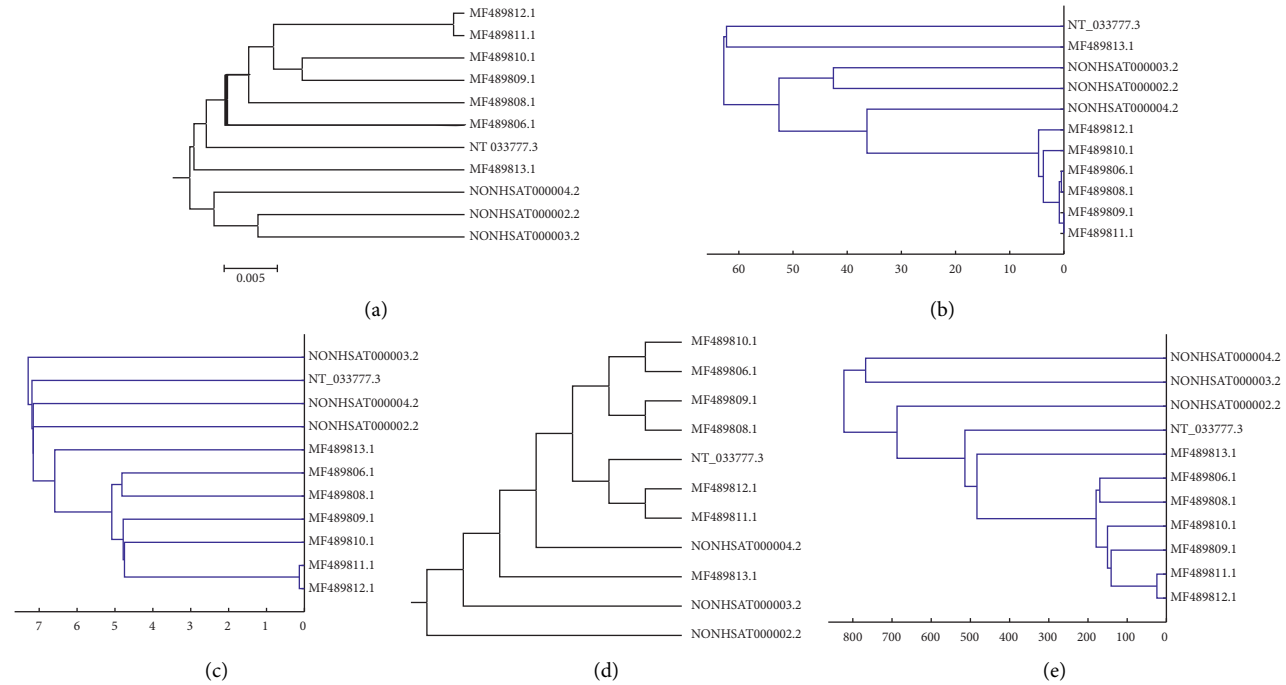


Figure 7: The cluster analysis graphs of 11 noncoding RNA sequences generated by (a) our method, (b) Li's method, (c) RNApdist, (d) ASPRAlign, and (e) MEGA software.

shown in Table 5. Clearly, the cluster analysis graph generated by our approach is closer to the cluster analysis graph generated by MEGA software than the other approaches.

In the second experiment, as indicated in Figure 7, the experimental sequences are accurately divided into three families in Figure 7(e), tRNA, RNase P RNA, and a group of human ncRNA in the NONCODE database. Figure 7(e) shows that the NT_033777.3 sequence has a high similarity to the tRNA family, which is also reflected in Figure 7(a) and Table 3. In Figure 7(b), NT_033777.3 and MF489813.1 are classified into a group, with NT_033777.3 being the most distant from the other sequences in the tRNA family, which is unreasonable. NT_033777.3 is mistakenly placed in a group close to the NONCODE human noncoding RNA in Figure 7(c). More unreliably, the grouping in Figure 7(d) does not completely separate the three groups of RNAs. On analysis of Table 4, it is apparent that ncRDeep did not succeed in accurately distinguishing ncRNAs. Based on the abovementioned analysis, the method based on DFA and DWT is an effective algorithm for RNA secondary structure similarity analysis. The distance between evolutionary trees generated by the method proposed by us, RNApdist, and Li's method, ASPRAlign, and the cluster analysis graph generated by MEGA software are shown in Table 6. Evidently, the results obtained by our method in this experiment are more accurate.

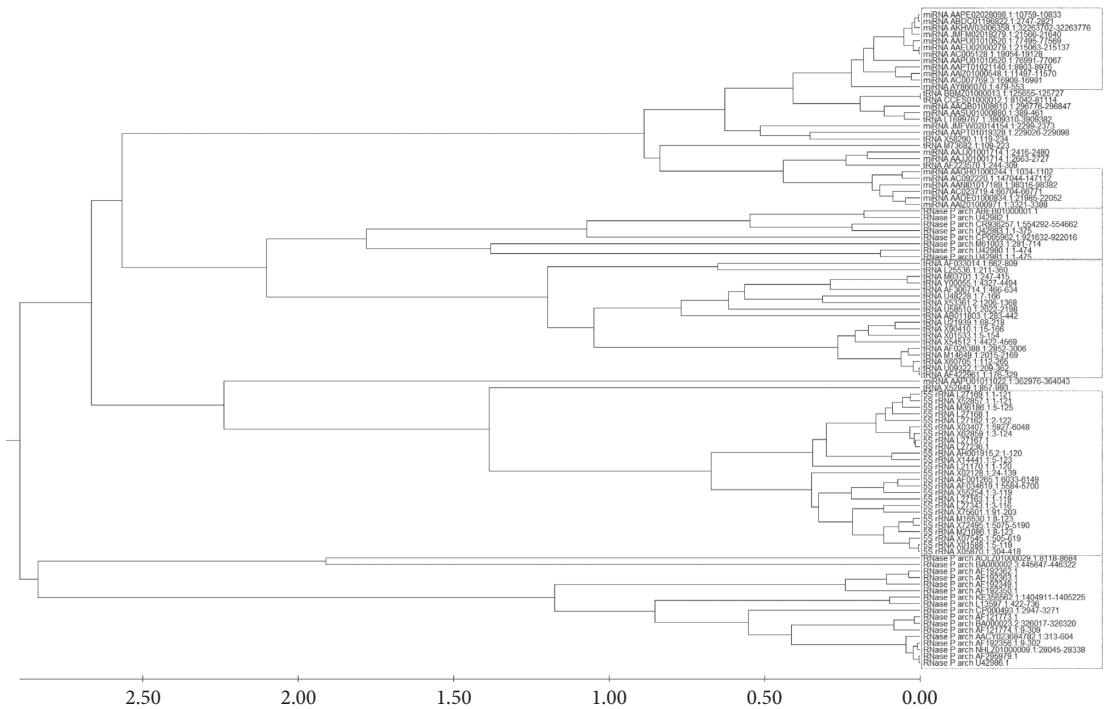


Figure 8: The cluster analysis graph of 100 noncoding RNA sequences generated by our method.

Table 5: The Pearson correlation efficiency between the cluster analysis graphs generated by the different methods and those generated by MEGA software.

Method	RNApdist	Our method	Li's method	ASPRAlign
Distance	0.1549	**0.1288**	0.2973	0.4709

Table 6: The Pearson correlation efficiency between the cluster analysis graphs generated by the different methods and those generated by MEGA software.

Method	RNApdist	Our method	Li's method	ASPRAlign
Distance	0.2494	**0.1748**	0.2048	0.2722

4. Conclusions

This paper proposes a hybrid method for the similarity comparison of RNA secondary structures. The algorithm is based on the existing RNA triple vector graphical representation, uses DWT to process the feature sequences, and captures the fractal characteristics using the DFA method. Compared with several commonly used RNA comparison methods, the approximate relationships between the RNA secondary structures obtained by the DFA and wavelet transform method are close to the actual relationships. However, the secondary structures predicted by the minimum free energy in the Vienna RNA package are not optimal, and finding the optimal secondary structure for RNA rapidly and efficiently remains a challenging problem. In addition, the method is not yet excellent for analyzing shorter RNAs, and systematic studies will be hoped to be carried out on RNAs of different lengths and characteristics.

Data Availability

The RNA sequence data used to support the findings of this study are deposited in the RFAM database.

Conflicts of Interest

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

Authors' Contributions

Lina Yang proposed the ideas and supervised the process of the research and the validity of the results. Yang Liu performed the practical experiments, recorded the experimental results, and wrote the manuscript. Xiaochun Hu guided the knowledge and application of the detrended fluctuation analysis (DFA) method. Patrick Wang provided guidance on the discrete wavelet transform. Xichun Li improved the language of the manuscript, and Jun Wu focused on the standardization of the manuscript format. All authors read and approved the final manuscript.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China under Grant No. 61862005.

Supplementary Materials

Supplementary Figure S1: The cluster analysis graph of 100 noncoding RNA sequences generated by Li's method. Supplementary Figure S2: The cluster analysis graph of 100 noncoding RNA sequences generated by *RNApdist*.

Supplementary Figure S3: The cluster analysis graph of 100 noncoding RNA sequences generated by *ASPRAlign*. (*Supplementary Materials*)

References

[1] L. R. Borges Dos Anjos, L. Do Prado Assunção, B. Lima Freitas et al., "Popularização da Ciência: desmistificando o Dogma Central da Biologia Molecular," *Revista de Ensino de Bioquímica*, vol. 16, no. 2, pp. 71–86, 2019.

[2] Y. Ke, R. Jiahua, H. Zhao, Y. Lu, N. Xiao, and Y. Yang, "Accurate prediction of genome-wide rna secondary structure profile based on extreme gradient boosting." 2019.

[3] W. Lu, Y. Tang, H. Wu et al., "Predicting RNA secondary structure via adaptive deep recurrent neural networks with energy-based filter," *BMC Bioinformatics*, vol. 20, no. 25, p. 684, 2019.

[4] T. Zhou and A. Routh, "Mapping RNA-capsid interactions and RNA secondary structure within virus particles using next-generation sequencing," *Nucleic Acids Research*, vol. 48, 2019.

[5] F. Wang, T. Akutsu, and T. Mori, "Comparison of pseudoknotted RNA secondary structures by topological centroid identification and tree edit distance," *Journal of Computational Biology*, vol. 55, 2020.

[6] J. H. Havgaard, E. Torarinsson, and J. Gorodkin, "Fast pairwise structural RNA alignments by pruning of the dynamical programming matrix," *PLOS Computational Biology*, vol. 3, no. 10, pp. 1896–1908, 2007.

[7] B. A. Shapiro and K. Zhang, "Comparing multiple RNA secondary structures using tree comparisons," *Bioinformatics*, vol. 6, no. 4, pp. 309–318, 1990.

[8] J. Allali and M. F. Sagot, *A New Distance for High Level RNA Secondary Structure Comparison*, IEEE Computer Society Press, Washington, DC,USA, 2005.

[9] D. Sankoff, "Simultaneous solution of the RNA folding, alignment and protosequence problems," *Siam Journal on Applied Mathematics-SIAMAM*, vol. 45, 1985.

[10] R. D. Dowell and S. R. J. B. B. Eddy, "Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints," *BMC Bioinformatics*, vol. 7, no. 1, pp. 1–18, 2006.

[11] D. H. Mathews, "Predicting a set of minimal free energy RNA secondary structures common to two sequences," *Bioinformatics*, vol. 21, no. 10, pp. 2246–2253, 2005.

[12] I. L. Hofacker, S. H. F. Bernhart, and P. F. Stadler, "Alignment of RNA base pairing probability matrices," *Bioinformatics*, vol. 20, no. 14, pp. 2222–2227, 2004.

[13] I. J. B. B. Holmes, "Accelerated probabilistic inference of RNA structure evolution," *BMC Bioinformatics*, vol. 6, 2005.

[14] H. J. Hull, G. D. Stormo, and G. J. B. Jan, "Pairwise local structural alignment of RNA sequences with sequence similarity less than," *Bioinformatics*, vol. 40, no. 9, p. 9, 2005.

[15] S. Will, K. Reiche, I. L. Hofacker, P. F. Stadler, and R. Backofen, "Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering," *PLOS Computational Biology*, vol. 3, no. 4, p. e65, 2007.

[16] S. Will, C. Schmiedl, M. Miladi, M. Möhl, and R. Backofen, "SPARSE: quadratic time simultaneous alignment and folding of RNAs without sequence-based heuristics," in *Proceedings of the 17th International Conference on Research in Computational Molecular Biology*, Santa Monica, USA, April 2013.

[17] S. Siebert and R. Backofen, "MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons," *Bioinformatics*, vol. 21, no. 16, pp. 3352–3359, 2005.

[18] E. Torarinsson, J. H. Havgaard, and J. Gorodkin, "Multiple structural alignment and clustering of RNA sequences," *Bioinformatics*, vol. 23, no. 8, pp. 926–932, 2007.

[19] H. Kiryu, Y. Tabei, T. Kin, and K. Asai, "Murlet: a practical multiple alignment tool for structural RNA sequences," *Bioinformatics*, vol. 23, no. 13, pp. 1588–1598, 2007.

[20] D. A. Sorescu, M. Möhl, M. Mann, R. Backofen, and S. Will, "CARNA--alignment of RNA structure ensembles," *Nucleic Acids Research*, vol. 40, no. W1, pp. W49–W53, 2012.

[21] C. B. Do, F. Chuan-Sheng, and B. J. B. Serafim, "A max-margin model for efficient simultaneous alignment and folding of RNA sequences," *Bioinformatics*, vol. 13, p. 13, 2008.

[22] R. D. Dowell, *RNA Structural Alignment Using Stochastic Context-Free Grammars*, Washington University, St. Louis, MO, USA, 2004.

[23] M. Hochsmann, T. Toller, R. Giegerich, and S. Kurtz, "Local similarity in RNA secondary structures," *Computational Systems Bioinformatics Csb IEEE Bioinformatics Conference Csb*, vol. 62, 2003.

[24] V. Guignon, C. Chauve, and S. Hamel, *An Edit Distance Between RNA Stem-Loops*, Springer, Berlin, Germany, 2005.

[25] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast folding and comparison of RNA secondary structures," *Monatshefte F and R Chemie Chemical Monthly*, vol. 125, no. 2, pp. 167–188, 1994.

[26] F. Jörg, V. Pavankumar, B. Andrea et al., "The RNA workbench 2.0: next generation rna data analysis," *Nucleic Acids Research*, vol. 47, 2019.

[27] M. Miladi, M. Raden, S. Will, and R. Backofen, "Fast and accurate structure probability estimation for simultaneous alignment and folding of RNAs," in *Proceedings of the 19th International Workshop on Algorithms in Bioinformatics*, Niagara Falls, NY, USA, September 2019.

[28] Q. Michela, T. Luca, and M. E. J. Bioinformatics, "ASPRAlign: a tool for the alignment of rna secondary structures with arbitrary pseudoknots," 2020.

[29] M. Quadrini, L. Tesei, and E. Merelli, "An algebraic language for RNA pseudoknots comparison," *Bmc Bioinformatics*, vol. 20, no. S4, 2019.

[30] L. Bo, C. Weiyang, S. Xingming et al., "A binary coding method of rna secondary structure and its application," *Journal of Computational Chemistry*, vol. 15, 2009.

[31] M. R. J. C. P. Letters, "On characterization of DNA primary sequences by a condensed matrix," *Journal of Chemistry*, vol. 317, no. 1, pp. 29–34, 2000.

[32] Y. H. Yao, B. Liao, T. M. J. Wang, and S. T. JoM, "A 2D graphical representation of RNA secondary structures and the analysis of similarity/dissimilarity based on it," *Journal of Molecular Structure*, vol. 755, no. 1, pp. 131–136, 2005.

[33] M. Randi and D. P. J. C. P. Letters, "Novel spectral representation of RNA secondary structure without loss of information," *Genomic Analysis and Structural Prediction of DNA*, vol. 476, no. 4–6, pp. 277–280, 2009.

[34] Y. Li, M. Duan, and Y. Liang, "Multi-scale RNA comparison based on RNA triple vector curve representation," *BMC Bioinformatics*, vol. 13, no. 1, p. 280, 2012.

[35] Y. Li, X. Shi, Y. Liang, J. Xie, Y. Zhang, and Q. Ma, "RNA-TVcurve: a Web server for RNA secondary structure comparison based on a multi-scale similarity of its triple vector curve representation," *BMC Bioinformatics*, vol. 18, no. 1, p. 51, 2017.

[36] M. Weiss and S. Ahnert, "Neutral components show a hierarchical community structure in the genotype–phenotype map of RNA secondary structure," *Journal of The Royal Society Interface*, vol. 17, 2020.

[37] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," 2020.

[38] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[39] J. Wu, S. Pan, X. Zhu, and Z. J. I. Cai, "Boosting for multi-graph classification," *IEEE Trans Cybern*, vol. 45, no. 3, pp. 416–429, 2014.

[40] T. Chantsalnyam, D. Y. Lim, H. Tayara, and K. T. Chong, "ncRDeep: non-coding RNA classification with convolutional neural network," *Computational Biology and Chemistry*, vol. 88, Article ID 107364, 2020.

[41] S. Siyu, M. A. Stephen, Y. Ruqiang et al., "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Transactions on Industrial Informatics*, vol. 63, no. 5, p. 1, 2019.

[42] S. Wang and X. Wang, "Prediction of protein structural classes by different feature expressions based on 2-D wavelet denoising and fusion," *BMC Bioinformatics*, vol. 20, no. Suppl 25, p. 701, 2019.

[43] J. Pando, L. Sands, S. E. J. P. R. E. S. N. Shaheen, and S. M. Physics, "Detection of protein secondary structures via the discrete wavelet transform," *Statistical, Nonlinear, and Soft Matter Physics*, vol. 80, no. 5, Article ID 051909, 2009.

[44] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Lilt, "A comparison of fractal dimension algorithms using synthetic and experimental data," *IEEE International Symposium on Circuits & Systems*, vol. 5, 1999.

[45] B. B. Mandelbrot and J. W. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[46] J. Li, X. Ma, M. Zhao, and X. J. E. Cheng, "A novel MFDFA algorithm and its application to analysis of harmonic," *Multifractal Features*, vol. 8, no. 2, 2019.

[47] Z. Yu, K. Lau, and L. Zhou, "Clustering of protein structures using hydrophobic free energy and solvent accessibility of proteins," *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 73, Article ID 031920, 2006.

[48] L. Yang, Y. Y. Tang, Y. Lu, and H. Luo, "A fractal dimension and wavelet transform based method for protein sequence similarity analysis," *IEEE/ACM Trans Comput Biol Bioinform*, vol. 12, no. 2, pp. 348–359, 2015.

[49] L. Yang, P. Wei, C. Zhong et al., "A fractal dimension and empirical mode decomposition-based method for protein sequence analysis," 2019.

[50] D. J. Kumar, C. P. Pal, C. Adwitiya, H. S. Sarif, and B. J. S. R. Pallab, "Analysis of purines and pyrimidines distribution over miRNAs of human, Gorilla, chimpanzee, mouse and rat," *Scientific Report*, vol. 8, no. 1, p. 9974, 2018.

[51] F. Hausdorff, "Dimension und *äβ*eres Ma*β*," *Mathematische Annalen*, vol. 79, no. 1, pp. 157–179, 1918.

[52] J. Michael, "Biology KJCi, Medicine," *Fractals and the Analysis of Waveforms*, vol. 18, no. 3, 1988.

[53] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," in *Proceedings of the IEEE Symposium on Computer-Based Medical Systems*, Lubbock, TX, USA, June 1995.

[54] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

[55] O. . Paweł, K. Jarosław, and E. D. J. P. R. Stanisław, "Wavelet versus detrended fluctuation analysis of multifractal structures," *Statistical, Nonlinear, and Soft Matter*, vol. 16, 2006.

[56] C.-K. Peng, S. Havlin, H. Stanley et al., "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," 1995.

[57] H. Jing, T. Wen-Wen, and G. Jianbo, "Detection of low observable targets within sea clutter by structure function based multifractal analysis," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 1, pp. 136–143, 2006.

[58] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[59] J. Alvarez-Ramirez, E. Rodriguez, I. Cervantes, J. Carlos Echeverria, and I. Applications, "Scaling properties of image textures: a detrending fluctuation analysis approach," *Physica A: Statistical Mechanics and Its Applications*, vol. 361, no. 2, pp. 677–698, 2006.

[60] C. Esser and S. Jaffard, "Divergence of wavelet series: a multifractal analysis," *Advances in Mathematics*, vol. 328, pp. 928–958, 2018.

[61] Z. Zhang, *Multivariate Wavelets. Multivariate Time Series Analysis in Climate and Environmental Research*, Springer International Publishing, Berlin, Germany, 2018.

[62] Z. Sun, Z. Zhang, Y. Chen, S. Liu, and Y. Song, "Frost filtering algorithm of SAR images with adaptive windowing and adaptive tuning factor," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 6, pp. 1097–1101, 2020.

[63] C. B. E. M. Reusken and J. F. Bol, "Structural elements of the 3′-terminal coat protein binding site in alfalfa mosaic virus RNAs," *Nucleic Acids Research*, vol. 24, no. 14, pp. 2660–2665, 1996.

[64] K. Sudhir, S. Glen, M. Li, K. Christina, and T. J. M. B. Koichiro, "Evolution. MEGA X," *Molecular Evolutionary Genetics Analysis Across Computing Platforms*, vol. 6, 2018.

[65] D. F. Robinson and L. R. Foulds, "Comparison of phylogenetic trees," *Mathematical Biosciences*, vol. 53, no. 1-2, pp. 131–147, 1981.

Hindawi
Complexity
Volume 2021, Article ID 6644111, 18 pages
https://doi.org/10.1155/2021/6644111

Research Article

Multiplex Network Embedding Model with High-Order Node Dependence

Nianwen Ning, Qiuyue Li, Kai Zhao, and Bin Wu

Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Beijing 100876, China

Correspondence should be addressed to Bin Wu; wubin@bupt.edu.cn

Received 9 October 2020; Revised 22 January 2021; Accepted 21 February 2021; Published 8 March 2021

Academic Editor: Shenghua Liu

Multiplex networks have been widely used in information diffusion, social networks, transport, and biology multiomics. They contain multiple types of relations between nodes, in which each type of the relation is intuitively modeled as one layer. In the real world, the formation of a type of relations may only depend on some attribute elements of nodes. Most existing multiplex network embedding methods only focus on intralayer and interlayer structural information while neglecting this dependence between node attributes and the topology of each layer. Attributes that are irrelevant to the network structure could affect the embedding quality of multiplex networks. To address this problem, we propose a novel multiplex network embedding model with high-order node dependence, called HMNE. HMNE simultaneously considers three properties: (1) intralayer high-order proximity of nodes, (2) interlayer dependence in respect of nodes, and (3) the dependence between node attributes and the topology of each layer. In the intralayer embedding phase, we present a symmetric graph convolution-deconvolution model to embed high-order proximity information as the intralayer embedding of nodes in an unsupervised manner. In the interlayer embedding phase, we estimate the local structural complementarity of nodes as an embedding constraint of interlayer dependence. Through these two phases, we can achieve the disentangled representation of node attributes, which can be treated as fined-grained semantic dependence on the topology of each layer. In the restructure phase of node attributes, we perform a linear fusion of attribute disentangled representations for each node as a reconstruction of original attributes. Extensive experiments have been conducted on six real-world networks. The experimental results demonstrate that the proposed model outperforms the state-of-the-art methods in cross-domain link prediction and shared community detection tasks.

1. Introduction

The abundant relations and views between entities can be collected from various sources or scenarios, allowing a slew of problems to be better solved in different application domains, e.g., information diffusion [1], social network analysis [2], intelligent transportation [3], biomedicine, and ecology [4, 5]. Taking together these data may be able to give a more accurate and nuanced picture of network structure than the individual network alone [6]. Taking social networks as an example, different online social networks show different views and behavior patterns of people. A user makes connections to their friends on Facebook or WeChat but uses Twitter or Weibo to follow people that interested him/her. Though different online social networks present distinct views and aspects of social behavior of one same user with the consistent feature, abundant user features and social information can facilitate the construction of a more accurate and nuanced user profile. Therefore, these multiple sources and views of network data are worth exploring because they often contain complementary information that improves the quality of analysis results [7].

Intuitively, modeling the information fusion problem of nodes as a feature fusion problem is a straightforward way. Based on the fused features, we can further mine the network data for node classification, link prediction, node clustering, and visualization. Multiple-relation or view network data are vividly modeled as a multiplex network (also known as multidimensional, multiview, or multilayer networks) [8–12] in which the same set of nodes are connected by

different types of relations. Different from a single network, multiplex networks reflect more complex topological properties. Multiplex networks can not only present the intralayer dependence between nodes but also can well model the interlayer network dependence. The analysis of multiple networks not only needs to consider the interdependence or interaction between nodes at the intralayer and interlayer but also focus on the dependence of node attributes and the topological structure of nodes. In this paper, the high-order node dependence of multiplex networks is defined as intralayer dependence between nodes, interlayer dependence in respect of anchor nodes, and the dependence between node attributes and the topology of each layer. In a multiplex network, the information fusion of multiple layers of nodes is a significant fundamental issue for the joint analysis of networks. A multiplex network, as shown in the middle of Figure 1, is composed of three social networks, which are Douban (https://www.douban.com/), LinkedIn (https://www.linkedin.com/), and Weibo (https://weibo.com/). These three social networks are geared towards different social scenarios; Douban provides books and music services, LinkedIn serves for social occupation, and Weibo is geared towards entertainment services. Multiplex network representation learning (also known as multiplex network embedding) is an effective method to analyze and mine the network. It can project the node (or network) into a continuous low-dimensional space. In this paper, we are motivated to focus on multiplex network representation learning considering the high-order dependence.

Recently, existing methods have achieved excellent performance in the intralayer dependence between nodes. However, few studies have comprehensively focused on the properties unique to multiplex networks. The first challenge is preserving high-order proximity information of nodes. Some state-of-the-art models based on the graph neural network (GNN) [10, 13, 14] take into account both intralayer and interlayer dependencies of nodes. However, due to the oversmoothing problem of GNN models [15], such methods cannot effectively preserve high-order proximity information. The second challenge is preserving the interlayer dependence property of multiplex networks. The layers with strong interlayer dependence have similar local structure characteristics, while those with weak interlayer dependence show obvious differences in the local topology [16]. From Figure 1, we can see that the nodes in the Douban layer and the Weibo layer have similar local structures. It indicates the interlayer dependence property of nodes in these two layers. This dependency cannot be preserved by extended random walk-based methods [17–19] and GNN-based methods [20–22]. The extended random walk-based representation learning method realizes the generation of node sequences through cross-layer sampling. In the node sampling process, most of them use random strategy to cross-layer sampling, but this ignores the similarity between layers. For GNN-based methods, nodes are embedded independently in interlayer. The node embedding of each layer is concatenated in the later stage. Such embedding and fusion processes will introduce repetitive and redundant information. However, LinkedIn layer is dissimilar with the other two layers. In this situation, it makes the fusion embedding of nodes obtained by methods [23–26] based on the assumption of information sharing between layers is inaccurate. The third challenge is preserving the dependence between node attributes and the topology of each layer. Previous studies also ignore the interaction of node attributes with the topology of each layer. Figure 1 illustrates this important property that different social scenarios depend on different attribute information of the user. The formation of friendship in the Douban network mainly depends on the user's preference for music and books. The formation of the following relationships in LinkedIn mainly depends on attributes such as the user's job and education level. The formation of relationships in Weibo mainly depends on the user's multiple attributes (books, sports, and music) besides job and education. In support of the dependence between the node attributes and the network structure, the interaction between them has been shown in several cases [27–29]. Therefore, the embedding of multiplex networks contains not only dependence information between nodes in each layer (intralayer dependence) but also local structure similarity information (interlayer dependence) and dependence between node attributes and the topology of each layer (attribute dependence).

In light of this, we propose a novel and hierarchy representation learning model for multiplex networks with node attributes called HMNE. We propose a symmetric graph convolution-deconvolution (GCD) method with multiple convolution layers to embed the intralayer adjacency information of a node as a low-dimensional dense vector in an unsupervised manner. The graph convolution module (GCM) preserves high-order proximity information, and the graph deconvolution module (GDM) serves as an embedding restriction to alleviate the oversmoothing problem of GCM. To preserve interlayer local dependence information, inspired by Graph Infomax [30], we use the similarity between the representations obtained by the multilayer convolution and the entire layer embedding as the estimation of complementary information. We fit this estimation of complementary information to actually quantify the local structural complementarity of nodes. For the dependence between attributes and the topology of the layer where the node is located, we treat the output of the graph deconvolution module as the disentangled representation of node attributes. Each disentangled representation is the result of the interaction between node attributes and the topology of each layer.

The main contributions of this paper are summarized as follows:

(i) We propose a symmetrical graph convolution-deconvolution neural network model to achieve intralayer node embedding, which is an unsupervised and general representation learning method. This method can not only flexibly adjust the number of hidden layers to capture the high-order structural information but also avoid the oversmoothing problem.

(ii) We present a method to estimate interlayer complementary information. This method can measure

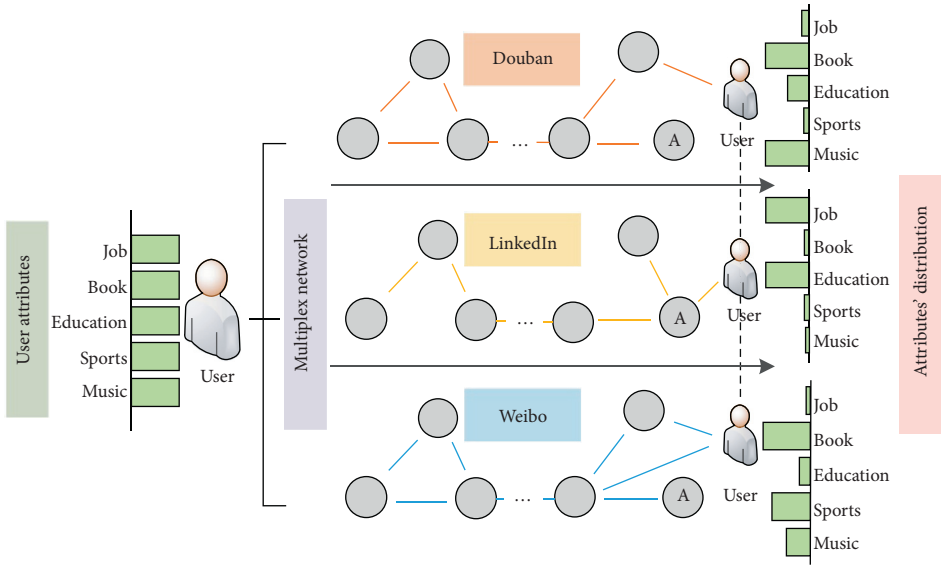


FIGURE 1: The illustration of our data structure and dependence between node attributes and each layer's topology for three-layer multiplex networks as an example. The first part on the left presents node attributes. The second part on the middle indicates the multiplex network data model. The latter part shows the different dependence between node attributes and each layer.

the interlayer dependence property [9, 10, 31] in respect of the topology of the layer where the node is located and constrain the intralayer embedding.

(iii) We design a disentangled representation learning architecture to solve the dependence between node attributes and its local topology. Graph deconvolution component is used to select attribute fragments associated with the semantics of each layer. We use a linear layer to restructure the original node attributes.

(iv) Extensive evaluations on real-world datasets have been conducted, and the experimental results demonstrate the superiority of the proposed HMNE model against the state-of-the-art models.

The rest of the paper is organized as follows. Section 2 describes some related works. Section 3 introduces related definitions of the data model we use, problem formulation, and preliminary knowledge. Section 4 presents HMNE's core modules. Section 5 shows the experiment results. Finally, the summary and outlook are described in Section 6.

2. Related Work

In this section, to distinguish from the single-layer network, we call the traditional representation learning method of one network as single-layer network embedding and the embedding of multiple networks as multiplex network embedding. Among them, we introduce the related work from joint embedding and cooperative embedding of multiplex network embedding. We first describe the ideas of network embedding for a single-layer network. Then, we, respectively, introduce related works about multiplex network (mainly involves multiview networks, multirelation networks, multidimensional networks, and multilayer networks) embedding methods. Finally, we also summarize the shortcomings of these related works and the similarities and dissimilarities with the proposed model.

2.1. Single-Layer Network Embedding

2.1.1. Random Walk-Based Methods. Embedding techniques based on random walk to obtain node representations have been proposed: DeepWalk [32] is the first algorithm based on random walk to learn node representation. Based on the breadth-first search and depth-first search, node2vec [33] was proposed to replace the node sampling strategy of the DeepWalk method. Both algorithms are traditional single-layer network embedding. Gu et al. [34] proposed an approach based on the open-flow network model to reveal the underlying flow structure and its hidden metric space of different random walk strategies on networks. It shows that the essence of network embedding by random walk is the latent metric defined on the open-flow network. In order to learn the representation of multirelation heterogeneous information networks, the following algorithm is proposed. Dong et al. [35] proposed a strategy for random walk sampling from heterogeneous networks, where the random walk is restricted to transition between particular types of nodes. This strategy allows many methods to be applied to heterogeneous graphs and complements the idea of taking type-specific encoders and decoders into account. Ribeiro et al. [36] presented struc2vec, a novel and flexible framework with the target to learn latent representations for the structural identity of nodes. The framework uses a hierarchy to measure node similarity at different scales and constructs a multilayer graph to encode structural similarities and generate a structural context of nodes.

2.1.2. Graph Neural Network-Based Methods. Kipf and Welling [37] introduced the variational graph autoencoder (VGAE), a framework for unsupervised learning on graph-

structured data based on the variational autoencoder (VAE). This model makes use of latent variables and is capable of learning interpretable latent representations for undirected graphs. Hamilton et al. [38] proposed GraphSAGE, which uses a two-layer deep neural architecture. In each convolution layer, a node computes its representation as an aggregation of its neighbors' representations (from the previous layer). In addition, to achieve unsupervised embedding, the parameters of aggregation functions are learned using the loss function similar to DeepWalk. GraphSAGE is incapable of selective neighbor sampling and has a lack of memory of known nodes that have been trained. To address these problems, Luo and Zhuo [39] proposed an unsupervised method that samples neighborhood information attended by co-occurring structures and optimizes a trainable global bias as a representation expectation for each node in the given graph. Velickovic et al. [30] presented Deep Graph Infomax (DGI), a general approach for learning node representations within graph-structured data in an unsupervised manner. DGI relies on maximizing mutual information between patch representations and corresponding high-level summaries of graphs, both derived using established graph convolution network architectures. Li et al. [29] proposed a principled unsupervised feature selection framework ADAPT to find informative features that can be used to regenerate the observed links and further characterize the adaptive neighborhood structure of the network. Yu et al. [15] proposed KS2L, a novel graph Knowledge distillation regularized Self-Supervised Learning framework, with two complementary regularization modules, for intra- and cross-model graph knowledge distillation. Xiao et al. proposed three rumor propagation models based on evolutionary game and antirumor [40], data enhancement [41], and representation learning [42]. They proved that rumors are not only influenced by antirumor information but also affected by user behavior and psychological factors. And they studied the user's network structure and historical behavior characteristics in the rumor topic communication space in social networks and predicted the user behavior in the next time slice based on the current time slice data. At the same time, they introduced evolutionary game theory and considered the internal and external factors that affect user behavior within rumor propagation.

2.2. Multiplex Network Embedding. The goal of multiplex network embedding methods is to achieve the information fusion of multiple features of networks, in which these methods can be divided into joint representation learning and coordinated representation learning [43] (in Figure 2 of [44], an illustration of coordinated and joint representation learning is presented).

2.2.1. Joint Representation Learning. Zhang et al. [24] proposed a scalable multiplex network embedding (MNE) method, which assumes that the same nodes in multiple networks preserve certain common features and unique features of each layer. Thus, the common and unique embedding of nodes in each layer is learned by the DeepWalk algorithm separately. Ma et al. [25] implemented node embedding for multidimensional networks with hierarchical structure. They simply added up node embedding in multiple dimensions as the fusion feature of nodes in multiple networks. Matsuno and Murata [26] presented a multilayer network embedding method (MELL) that captures and characterizes each layer's connectivity. The method utilizes the overall structure to consider the similar or complementary structure of the layer. Finally, the fusion feature learning of nodes in multiplex networks is obtained by combining node embedding in each layer with layer vectors. Cen et al. [9] focused on embedding learning for attributed multiplex heterogeneous networks, where different types of nodes might be linked with multiple different types of edges, and each node is associated with a set of different attributes. GATNE splits the overall node embedding into three parts: base embedding, edge embedding, and attribute embedding. GATNE-T contains only the first two parts. Zhao et al. [45] proposed a novel and principled approach: a multiview adversarial completion model (MV-ACM). Each relation space is characterized in a single viewpoint, enabling us to use the topological structural information in each view. Yuan et al. [46] proposed a novel multiview network embedding model with node similarity ensembles. Node similarities are first selected to maximize the represented network information while minimizing the information redundancy. For each combination of the selected node similarities, a latent space is generated as a view of the network.

2.2.2. Coordinated Representation Learning. In some cases, graphs have multiple "layers" that contain copies of the same nodes. They can be beneficial to share information across layers so that a node's embedding in one layer can be informed by its embedding in other layers. Qu et al. [47] proposed an attention-based method (MVE) to learn the weights of views for different nodes with a few labeled data. MVE can obtain robust node representations across different views by vote strategy. Recently, Liu et al. [17] extended a standard graph mining into the area of the multilayer network. The proposed methods ("network aggregation," "results' aggregation," and "layer coanalysis") can project a multilayer network of a continuous vector space. Zitnik and Leskovec [18] proposed the OhmNet framework to learn the features of proteins in different tissues. They represented each tissue as a network, where nodes represent proteins. Individual tissue networks act as layers in a multilayer network, where they use a hierarchy to model dependencies between the layers (i.e., tissues). Schlichtkrull et al. [20] introduced relational graph convolution networks (R-GCNs) and applied them to two standard knowledge base completion tasks: link prediction (recovery of missing facts, i.e., subject-predicate-object triples) and entity classification (recovery of missing entity attributes). Zhiyuli et al. [48] proposed highly scalable node embedding for link prediction in large-scale networks. The method learns node pairs' co-occurrence features to embed a node into a vector by a damping-based random walk

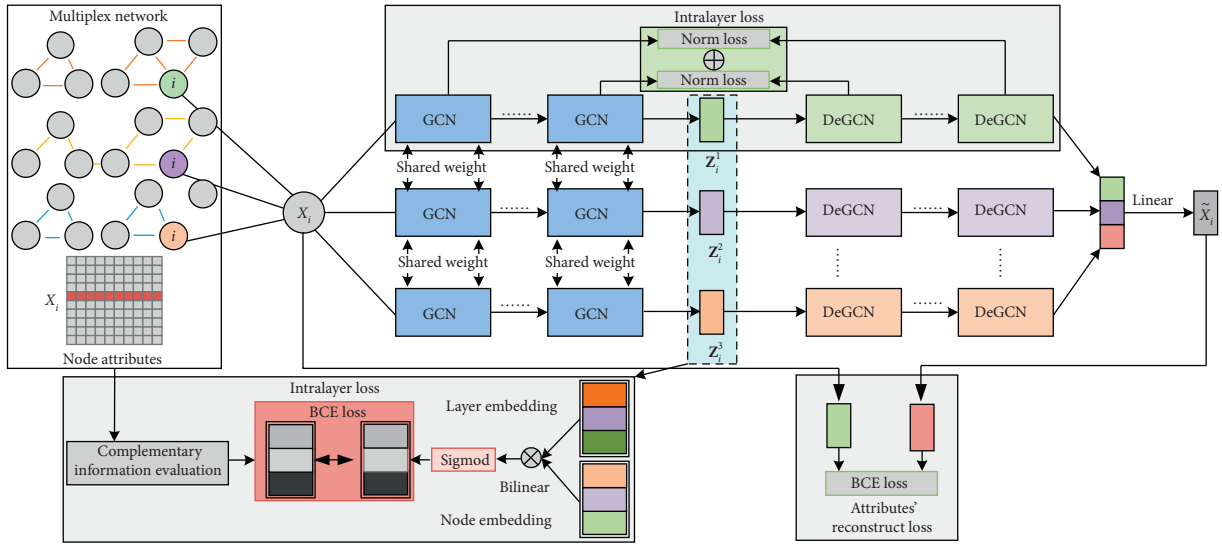


Figure 2: The architecture of HMNE. Our model includes the calculation of three types of loss functions, namely, intralayer embedding loss with shared weight, interlayer embedding loss, and global reconstruction loss. In this figure, GCN is a graph convolution neural layer, and DeGCN is a graph deconvolution neural layer.

algorithm. In the node sampling process, there is a bias problem with these existing methods that samples are trapped in a local structure. In addition, cross-layer sampling heavily depends on fixed parameters, which is in an inflexible manner. Sun et al. [49] presented a MNGAN framework for multiview network embedding by the generative adversarial network, aimed at preserving the information from the individual network views while accounting for connectivity across different views. Wei et al. [50] proposed an attributed node random walk framework, which can not only be able to incorporate both topology and attribute information flexibly but also easily deal with missing data and is applied to large networks. For the multiple-network alignment problem, Chu et al. [51] proposed a cross-network embedding method (CrossMNA). It defines two categories of embedding vectors for each node: intervector, and intravector. The idea of CrossMNA is the same as that of MNE. They thought intravector contains both the commonness among counterparts and the specific local connections in its selected network due to the semantics. Park et al. [10] presented a simple yet effective unsupervised network embedding method for the attributed multiplex network called DMGI, inspired by Deep Graph Infomax (DGI), which maximizes the mutual information between local patches of a graph and the global representation of the entire graph. Vashishth et al. [21] proposed a novel graph convolutional framework (COMPGCN) which jointly embeds both nodes and relations in a relational graph. COMPGCN leverages a variety of entity-relation composition operations from knowledge graph embedding techniques and scales with the number of relations. Yu et al. [22] proposed a novel GEneralized Multirelational Graph Convolutional Networks framework, which combines the power of GCNs in graph-based belief propagation and the strengths of advanced knowledge-based embedding methods, and goes beyond.

In summary, in response to the challenges presented in this paper, the single-layer network embedding methods cannot achieve the preservation of interlayer-dependent information. The joint representation learning methods of multiplex networks assume that nodes have shared embeddings in interlayer, and information sharing and transfer are realized through these embeddings. However, different levels of dependence between layers will cause this assumption to be invalid (please refer to Figure 3 of literature [44]). The existing coordinated representation learning methods neglect node attributes and their local topology. Aggregating these coarse-grained attributes in the graph neural network can include noise and affect the performance of the model. In order to fill this gap, we propose a hierarchical multiplex network embedding (HMNE) model with high-order node dependence. The specific implementation will be described in detail in Section 4.

3. Data and Problem Formulations

In this section, we describe related symbols, concepts, and definitions in detail. Our data model's basic concepts are introduced in Section 3.1. Then, we formalize a generalized node embedding problem of multiplex networks in Section 3.2. The important notations are summarized in Table 1.

3.1. Data Model. In terms of network data of multiple views and sources, it is more appropriate to represent such networks as multiplex networks. As shown in Figure 1, three layers of this multiplex network are derived from three modal data, such as social network, semantic relation network, and co-occurrence network. Multiplex networks can not only express the intralayer link but also can well model the dependencies and interactions between networks [44]. The detailed definitions of multiplex networks are as follows.

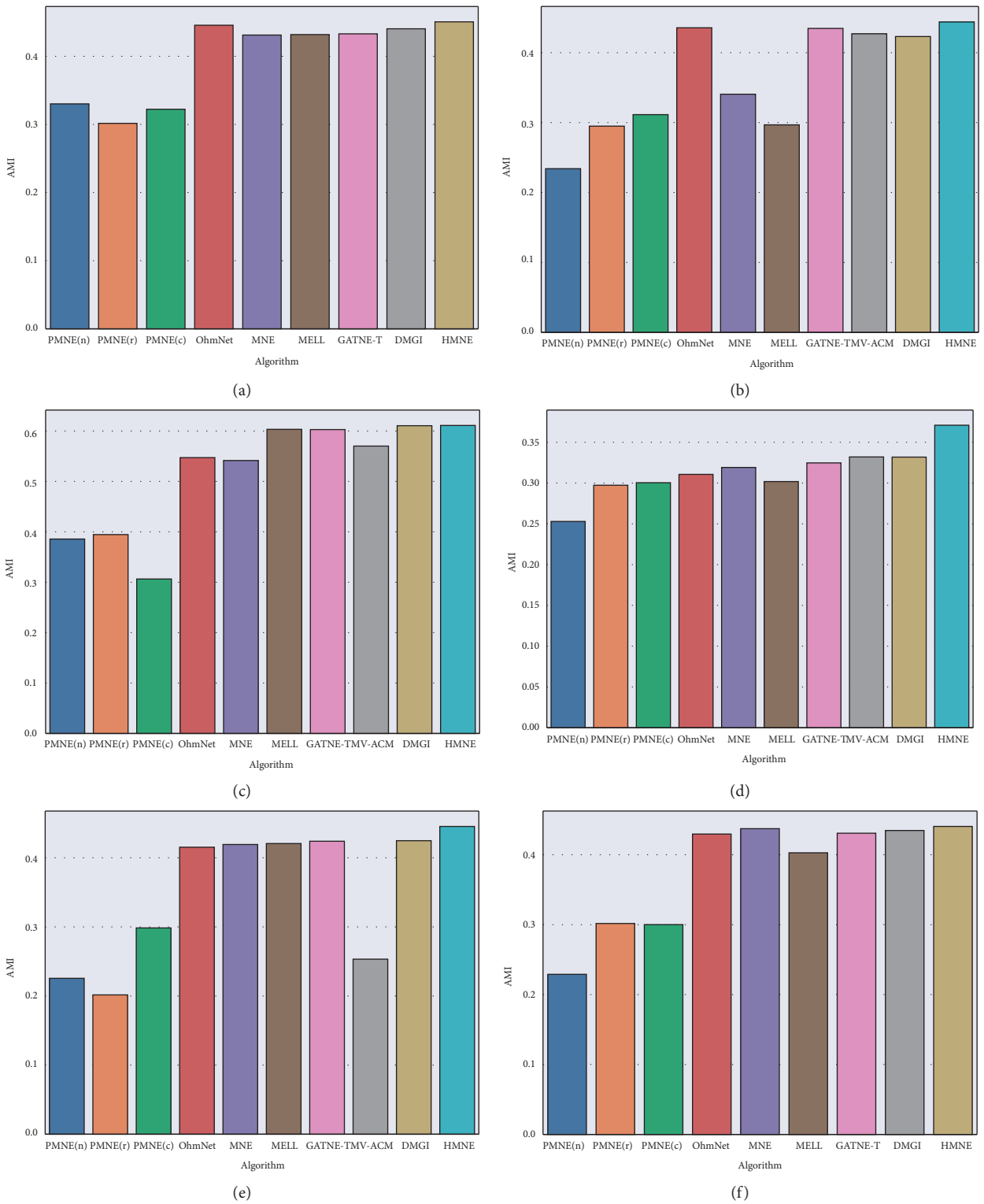


Figure 3: The AMI scores of HMNE in real-world multiplex networks. For the cross-domain link prediction task, AMI is regarded as an indicator to evaluate the embedding performance. (a) Celegans. (b) CKM. (c) CS-Aarhus. (d) London. (e) Vickers. (f) Ack-co-author.

Definition 1 (multiplex network architecture). Given a multiplex network of N nodes with the sets of layer L, in which each node can interact with the other ones through $|L|$ kinds of relations with $|L| \geq 2$, we denote an aligned multiplex network $G = \{G^l(\mathscr{V}, \mathscr{E}^l), l \in L\}$ which is made up of $|L|$ layers with $N = |\mathscr{V}|$ nodes and $E = |\sum_{l \in L} \mathscr{E}^l|$ edges.

Table 1: Notations.

Notation	Explanation
G	A multiplex network
$\mathcal{V}, \mathcal{E}^{\alpha}$	The sets of nodes and edges in layer α, respectively
G^l, A^l	A network/adjacency matrix of layer l, respectively
N, E^l	The node number/edge number of layer l, respectively
$\mathcal{X}$	$\{x_i \mid i \in \mathcal{V}\}$, the set of node attributes
$\mathcal{H}$	The node representations of multiplex networks
i, x_i	A node and its features, respectively
$\deg(i)$	The degree of node i
$e^l_{i,j}$	An edge between i and j in layer l
k	The number of neural network layers
$\mathcal{N}^l_i$	Neighbors of node i in layer l
$\mathbf{h}^l_i$	The learned representation of node i in layer l
d	The dimension of node representations
$\mathbf{Z}^l, \widetilde{\mathbf{Z}}^l$	The output of GCC/GDC in layer l of G
$\mathbf{Z}^{(k)}, \widetilde{\mathbf{Z}}^{(k)}$	The output of the k-th convolution/deconvolution layer
Θ, Θ_d	The convolution/deconvolution kernel

Each layer in multiplex networks has the same node set and different edge sets, as shown in the middle part of Figure 1. Let $i, j \in \mathcal{V}$ be two nodes. i^l denotes node i at layer l, and $e^l_{i,j} \in \mathcal{E}^l$ denotes the edge to link i^l and j^l in layer l. i^l and $i^{l'}$ are the duplicates of the same node i in different layers. We assume that nodes i^l and $j^{l'}$ can be implicitly linked by the duplicates of i in layer l' and $e^{l'}_{i,j}$ cross-layers l and l'. Figure 1 shows an illustrative example of a multiplex network with $|L| = 3$-layer network (i.e., $L = \{\text{Douban, LinkedIn, Weibo}\}$) and a target node User. The dotted line represents an anchor link. $e^{\text{LinkedIn}}_{\text{User},A}$ is an edge between node User and node A in layer LinkedIn. $e^{\text{Douban,LinkedIn}}_{\text{User},A}$ is a cross-layer link between node $\text{User}^{\text{Douban}}$ and node A^{LinkedIn} through an anchor link.

3.2. Problem Formulation

Definition 2 (multiplex network representation learning). Suppose the methods make use of a real-valued superadjacency matrix A, $A \in R^{(N\times|L|,N\times|L|)}$ (e.g., representing text or metadata associated with nodes). Node embedding aims at learning a map function f: $A \longrightarrow \mathcal{H}$.

f is a function, which maps $A_i = \{A_i^{l_1}, A_i^{l_2}, \ldots, A_i^{l_{|L|}}\} \in R^{(N,|L|)}$ to a d-dimensional representation of node i, and A_i is a group of vectors of node i in the superadjacency matrix of G, and it can also be understood that it is composed of adjacency matrices of multiple layers. $\mathcal{H}$ is a d-dimensional vector/tensor, and $d \ll N$. For coordinated representation learning, $\mathbf{h}_i$ is a vector for node i. For joint representation learning, $\mathbf{h}_i$ is a tensor for node i. Notice that all the aforementioned definitions can be easily extended to the case of weighted networks. We only focus on coordinated representation learning in this paper.

4. Proposed Model

In this section, we introduce the overall model of our HMNE by addressing the three major challenges mentioned in Section 1:

(1) Preserving high-order proximity information of nodes: as shown in Figure 2, a symmetric graph convolution-deconvolution network (SGCD) model is designed to solve the oversmoothing problem of the traditional GCN. GCD includes the graph convolution component (GCC) and graph deconvolution component (GDC). We formulate a restriction constraint for the GDC to restructure the original input feature of the GCC. The output feature of the GCC with K (graph) convolution layers $\mathbf{x}_i^k$ in respect of node i is inputted into the GDC for reconstructing original input feature x_i. Even if many graph convolution layers are added to the GCC, the oversmoothing problem can be avoided because of this reconstruction constraint. Therefore, we can conveniently preserve high-order proximity information of nodes by increasing the graph convolution layers.

(2) Preserving the interlayer dependence property of multiplex networks: as shown in Figure 2, there are two major components to capture the intralayer dependence property of multiplex networks. We first utilize a structural similarity metric method to measure the difference target layer l and the other layer l', respectively, in respect of node i. The result is served as a structural complementary information estimation P_{true}. Then, the similarity measure between the embedding $\mathbf{h}_i^l$ of node i in the target layer l and the global embedding $\mathbf{H}^{l'}$ in the other layer l' is served as the complementary information P_{pred} in respect of node i. Through the minimization of P_{pred} and P_{true}, the learned embedding of node i can preserve the dependency property between layers.

(3) Preserving the dependence of node attributes with the topology of each layer: as shown in Figure 2, the input feature is attributes x_i of node i. We utilize the idea of disentanglement learning to disentangle x_i as $|L|$ attribute subsets. These attribute subsets dependent on the topology of each layer have different semantic information. Three main processes are as follows: firstly, we use x_i of node i as the input of the GCC. Then, the embedding $\mathbf{h}_i^l$ of node i with attribute information and structure information is obtained by the GCC in layer l of multiplex networks. Finally, the disentangled representations of the node's attributes are the output of the GDC in each layer. In the GCC, the attributes associated with the topology of each layer are preserved. In the GDC, the structure information is disentangled from $\mathbf{h}_i^l$.

4.1. Preserving High-Order Proximity

4.1.1. Graph Convolution. In spectral-based graph convolution models, a mathematical representation of an undirected graph is the normalized graph Laplacian matrix defined as $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-(1/2)}A\mathbf{D}^{-(1/2)}$, where $\mathbf{D}$ is a diagonal matrix of node degrees. The normalized Laplacian matrix can be factored as $\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^{\mathbf{T}}$, where Λ is the diagonal matrix of eigenvalues. The eigenvectors of the normalized Laplacian matrix form an orthonormal space; in mathematical words, $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. In graph signal processing, a feature vector of node i of a graph is a graph signal $x_i \in \mathbf{R}^N$.

The graph Fourier transform to a signal x_i is defined as $\mathscr{F}(x_i) = \mathbf{U}^T x_i$, and the inverse graph Fourier transform is defined as $\mathscr{F}^{-1}(\widehat{x}_i) = \mathbf{U}\widehat{x}_i$, where $\widehat{x}_i$ represents the resulting signal from the graph Fourier transform. The graph convolution of the input signal x_i with a convolution kernel (filter) $\mathbf{g}$ is defined as

$$x_i *_G \mathbf{g} = \mathscr{F}^{-1}\left(\mathscr{F}(x_i) \odot \mathscr{F}(\mathbf{g})\right) = \mathbf{U}\left(\mathbf{U}^T\mathbf{x} \odot \mathbf{U}^T\mathbf{g}\right), \tag{1}$$

where $\odot$ denotes the Hadamard product. If we denote a filter as $\mathbf{g}_\theta = \mathrm{diag}(\mathbf{U}^T\mathbf{g})$, then the graph convolution is simplified as

$$x_i *_G \mathbf{g} = \mathbf{U}\mathbf{g}_\theta\mathbf{U}^T x_i. \tag{2}$$

The graph convolution component from [52] limits the layerwise convolution operation to alleviate the problem of overfitting on local neighborhood structures for graphs with very wide node degree distributions. The equation simplifies to

$$x_i *_G \mathbf{g} \approx \theta_0' x_i + \theta_1'(\mathbf{L} - \mathbf{I}_N)x_i = \theta_0' x_i - \theta_1'\mathbf{D}^{-(1/2)}\mathbf{A}\mathbf{D}^{-(1/2)}x_i. \tag{3}$$

After constraining the number of parameters with $\theta = \theta_0' = -\theta_1'$, we can obtain the following expression:

$$x_i *_G \mathbf{g} \approx \theta\left(\mathbf{I}_N + \mathbf{D}^{-(1/2)}\mathbf{A}\mathbf{D}^{-(1/2)}\right)x_i. \tag{4}$$

Kipf and Welling [52] introduced the trick: $\mathbf{I}_N + \mathbf{D}^{-(1/2)}\mathbf{A}\mathbf{D}^{-(1/2)} \approx \widetilde{\mathbf{D}}^{-(1/2)}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-(1/2)}$, where $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ and $\widetilde{\mathbf{D}}_{ii} = \sum_j \widetilde{\mathbf{A}}_{ij}$. Finally, we treat Θ as a convolution kernel (a matrix of filter parameters), a general definition of graph convolution as follows:

$$\mathbf{Z} = \widetilde{\mathbf{D}}^{-(1/2)}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-(1/2)}\mathbf{X}\boldsymbol{\Theta}. \tag{5}$$

In order to express the following sections more clearly, we denote $\mathbf{Z}^l$ as the node embedding of layer l of a multiplex network G and $\mathbf{Z}^{(k)}$ as a node embedding output of the k-th layer of the graph convolution neural network.

4.1.2. Graph Deconvolution. To capture the high-order proximity information of the nodes, we can simply stack multiple convolution layers as our HMNE's graph convolution component (GCC) based on equation (5). However, previous studies showed that graph convolution is a type of Laplacian smoothing. They proved that, after repeatedly applying Laplacian smoothing many times, the features of the nodes in the (connected) graph would converge to similar values. To avoid this problem and capture the high-order proximity of nodes, we design a graph deconvolution component (GDC). We first take the output $\mathbf{Z}^{(k)}$ of the k (multiple) stacked convolution neural layers as the input of the GDC. Then, analogous to the definition of deconvolution in the field of computer vision, according to equation (5), a graph deconvolution layer with a deconvolution kernel Θ_d is defined as

$$\widetilde{\mathbf{Z}} = \widetilde{\mathbf{D}}^{-(1/2)}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{(1/2)}\mathbf{Z}^{(k)}\Theta_d, \tag{6}$$

where A is an adjacency matrix, $A \in R^{N\times N}$, $\widetilde{A} = A + I_N\widetilde{D}$ is a degree matrix, and $\widetilde{D}_{ii} = \sum_j A_{ij}$. The embedding of nodes $\widetilde{\mathbf{Z}}^{(k)}$ is an output of the k-th layer of the graph deconvolution neural network.

4.1.3. Intralayer Embedding Loss. In this initialization of the GDC, the input matrix $\widetilde{\mathbf{Z}}^{(1)}$ is $\mathbf{Z}^{(k)}$, where k is the number of graph convolution layers, and $\mathbf{Z}^{(k)}$ is a final node embedding matrix according to equation (5). To separate the structure information from the input by the deconvolution kernel, we propose an intralayer embedding loss formula. We use a symmetric structure containing k convolution layers and k deconvolution layers as our graph convolution-deconvolution component (SGCD). The reconstruction loss formula of SGCD is

$$L_{\text{intra}} = \sum_{j=1}^{(K/2)} \left\| \widetilde{\mathbf{Z}}^{(j)} - \mathbf{Z}^{(K-j)} \right\|_2^2. \tag{7}$$

We assume the input $\mathbf{Z}^{(1)}$ of the GCC is the node attributes $\mathscr{X}$ so that the output $\widetilde{\mathbf{Z}}^{(k)}$ of the GDC is a reconstruction matrix in respect of $\mathscr{X}$. This reconstruction process is significant for Section 4.3.

4.1.4. Node Representation Learning. In order to preserve the attribute and structural information of the node in each layer, we need to aggregate the embedding $\mathbf{h}_i^l$ ($l \in L$) of node i in each layer to obtain a more complementary global node embedding $\mathbf{h}_i$. We use a sum function to integrate the embeddings of node i in each layer:

$$\mathbf{h}_i = \sum_{l\in L} \mathbf{h}_i^l. \tag{8}$$

Then, the final embedding of nodes in multiplex networks is

$$\mathscr{H} = \sum_{l\in L} \mathbf{Z}^l. \tag{9}$$

The final embedding $\mathbf{h}_i^l \in \mathbf{Z}^l$ of node i in layer l obtained by the GCC, where $\mathbf{Z}^l \in \mathscr{R}^{N\times d}$ and $\mathbf{h}_i^l \in \mathscr{R}^{1\times d}$, is a row of $\mathbf{Z}^l$.

4.2. Preserving Interlayer Dependence

4.2.1. Interlayer Structure Complementary Information Estimation. To capture the interlayer dependence between

layers, we introduce how to get the true sample (l', l, i), which indicates layer l' is complementary for i in l. This complementary information computing can effectively measure the interlayer dependence property. The basic idea is that the more dissimilar the local structures in two layers, the more reason to believe complementary information exists between these two layers. So, we utilize the structural similarity between two layers to produce true samples. Let $P_{\text{true}}(\cdot \mid i, l)$ denote the true underlying connecting distribution of node i in layer l, and we can estimate it as

$$p(j \mid i, l) = \frac{e_{i,j}^{l}}{\sum_{v \in \mathscr{V}} e_{i,v}^{l}}. \tag{10}$$

Then, the locally topological structural similarity of node i between layers l' and l can be calculated by Jensen–Shannon distance between $P_{l',i} = P(\cdot \mid i, l')$ and $P_{l,i} = P(\cdot \mid i, l)$ as

$$D_{\text{JS}}\left(P_{l',i} \| P_{l,i}\right) = \frac{1}{2}\left[D_{\text{KL}}\left(P_{l,i} \| M\right) + D_{\text{KL}}\left(P_{l',i} \| M\right)\right], \tag{11}$$

where $M = ((P_{l,i} + P_{l',i})/2)$ and D_{KL} is the Kullback–Leibler divergence:

$$D_{\text{KL}}(P \| Q) = \sum_{v} P(v) \log \frac{P(v)}{Q(v)}. \tag{12}$$

Note that when the locally topological structures of node i between layers l' and l are identical, $D_{\text{JS}}(P_{l',i} \| P_{l,i}) = 0$; otherwise, $D_{\text{JS}}(P_{l',i} \| P_{l,i}) = 1$. So, we get $S_{\text{struc}}(l', l \mid i) = 1 - D_{\text{JS}}(P_{l',i} \| P_{l,i})$ as the locally topological structural similarity between layers l' and l regarding node i. Finally, we can estimate $P_{\text{true}}(\cdot \mid l, i)$ and sample true layers according to the distribution:

$$P_{\text{true}}(\cdot \mid l, i) = \underset{l' \in L}{\Delta} \frac{S_{\text{struc}}(l', l \mid i)}{\sum_{r \in L} S_{\text{struc}}(r, l \mid i)}, \quad l' \in L, \tag{13}$$

where Δ denotes a function that can concatenate each element successively. Actually, this structure complementary information estimation can be served as the similarity of node i in layer l with respect to the topology of the layer where the node is located.

4.2.2. The Interlayer Dependence Estimation of Nodes. In order to realize the interlayer dependence property, inspired by the idea of Deep Infomax in [53], we regard the membership of node i in layer l for layer l' as a measure of the interlayer local dependency of node i. Therefore, a layer-level embedding $\mathscr{H}^{l}$ of layer l in multiplex networks is computed by employing a readout function Readout: $R^{n \times d} \longrightarrow \mathscr{R}^{d}$.

$$\mathscr{H}^{l} = \text{Readout}\left(\mathbf{Z}^{l}\right) = \sigma\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{h}_{i}^{l}\right), \tag{14}$$

where $\mathbf{Z}^{l}$ is a final embedding matrix of layer l in the graph convolution component, $\mathbf{h}_{i}^{l}$ is an embedding of node i of the l layer, and σ is a logistic sigmoid nonlinearity function.

Based on the layer-level embedding and the embedding of each node in this layer, we calculate the measure of the interlayer dependence property of node i in layer l on layer l'. In this paper, we apply a simple bilinear scoring function as it empirically performs the best in our experiments:

$$\text{Score}(l' \mid l, i) = \text{Score}\left(\mathbf{h}_{i}^{l}, \mathscr{H}^{l'}\right) = \sigma\left(\mathbf{h}_{i}^{l} \mathbf{W} \mathscr{H}^{l'}\right), \tag{15}$$

where σ is the logistic sigmoid nonlinearity and $\mathbf{W} \in \mathscr{R}^{d \times d}$ is a trainable scoring matrix. We can estimate the interlayer local dependence measure of the nodes by calculating the scores of the nodes' embedding in each layer and the global embedding of each layer:

$$P_{\text{pred}}(\cdot \mid l, i) = \underset{l' \in L}{\Delta} \text{Score}(l' \mid l, i), \tag{16}$$

where $P_{\text{pred}}(\cdot \mid l, i)$ denotes a vector of interlayer dependence of node i in layer l in respect of the duplication of node i in each layer and Δ denotes a function that can concatenate each element successively.

4.2.3. Interlayer Dependence Loss. Comparing equation (13) with (16), we have designed an objective function with BCELoss loss function for saving the node interlayer dependence property:

$$\begin{aligned} L_{\text{inter}} = \frac{1}{N}\sum_{i \in \mathscr{V}}\sum_{l'=1}^{L} -\Big[& P_{\text{true}}(l' \mid l, i) \log \widetilde{P}_{\text{pred}}(l' \mid l, i) \\ & + (1 - P_{\text{true}}(l' \mid l, i)) \log\left(1 - \widetilde{P}_{\text{pred}}(l' \mid l, i)\right)\Big]. \end{aligned} \tag{17}$$

4.3. Preserving Dependence between Attributes and Topology. In order to preserve the dependence between attributes and the topology of each layer, the original attributes of nodes are fed into the GCC. We perform GCC and GDC processes to disentangle the attributes of nodes as different semantic representations. We believe that the GCC can strengthen the attribute value related to the layer's semantic in the node attributes. GDC can disentangle the attributes of nodes with structure information of nodes. This is the main advantage of our GCD (intralayer embedding) method compared with the graph autoencoder and variational autoencoder. Then, in the final graph deconvolution network phase, each final output embedding of the GDC for each layer of multiplex networks is aggregated by a concatenate function. A linear layer is used to reconstruct the original attributes x_i, which makes the overall model framework designed as an autoencoder architecture. Based on the embedding of node i in layer l and the embedding of node i in other layers, we construct a simple nonlinear fusion method to obtain the reconstruction attributes $\widetilde{x}_i$ of node i:

$$\widetilde{x}_i = \sigma\left(W \Delta_{l'=1}^{L} \widetilde{\mathbf{Z}}_{i}^{l'}\right), \tag{18}$$

where σ is a sigmoid nonlinearity activation function, W is the trainable parameters, and $\widetilde{\mathbf{Z}}_{i}^{l'}$ is the output of the GDC of i node in the l' layer network. Then, we also utilize the

BCELoss function to calculate the loss between original attributes x_i and reconstruction attributes $\tilde{x}_i$ of node i:

$$L_{\text{attr}} = \frac{1}{N} \sum_{i \in \mathcal{V}} \sum_{l \in L} -\left(\left(x_i \log \tilde{x}_i^l + (1 - x_i) \log\left(1 - \tilde{x}_i^l\right)\right)\right), \quad (19)$$

where L is the layer number of multiplex networks and x_i is the attributes of the i node.

Finally, the global loss function of HMNE also considers the loss of different components. Therefore, we simply sum all the loss functions as the loss of the entire model and use Adam optimizer for backpropagation and parameter learning. The loss function of HMNE is

$$L = L_{\text{inter}} + L_{\text{intra}} + L_{\text{attr}}. \quad (20)$$

4.4. The Optimization and Time Complexity. We present the node representation learning process (HMNE) for multiplex networks in Algorithm 1. The total time complexity of HMNE is $O(\text{TNE}|L|^2)$ where T is the number of iterations, N is the number of nodes in each layer, E is the number of edges of the multiplex network, and $|L|$ is the number of layers.

5. Experiment Analysis

In this section, we study the performance of HMNE in different real-world datasets. We use cross-domain link prediction and shared community detection tasks to verify the performance of HMNE.

5.1. Datasets. For our experiments, we conduct HMNE and compare baseline methods on each of the following multiplex networks. These datasets contain two categories: public datasets and private dataset. Public datasets are composed of five multiplex network benchmark datasets involving social, biological, and transportation. Private dataset is an interesting semantic network dataset that we construct. This dataset is a network of acknowledgment relationships extracted from the acknowledgment part of dissertation data and the coauthor network of corresponding entities from AMiner (https://www.aminer.cn/). The specific information about public and private datasets is shown in Table 2.

5.1.1. Public Datasets. These multinetwork datasets were collected on M. De Domenico's homepage (https://comunelab.fbk.eu/manlio/index.php), and the processed datasets are available (https://github.com/Brian-ning/HMNE/).

Vickers classroom social multiplex network: this dataset was collected by Vickers from 29 seventh-grade students in a school in Victoria, Australia. Students were asked to nominate their classmates on a number of relations (class, best friend, and work).

CS-Aarhus social multiplex network: this dataset consists of five kinds of online and offline relationships (Facebook, leisure, work, coauthorship, and lunch) between the employees of the computer science department at Aarhus. These variables cover different types of relations between the actors based on their interactions.

London multiplex transport network: this dataset was collected in 2013 from the official website of Transport for London and manually cross-checked. Nodes are train stations in London, and edges encode existing routes between stations. Tube, overground, and DLR stations are considered.

CKM physicians' innovation multiplex network: this dataset was collected by Coleman, Katz, and Menzel on medical innovation, considering physicians in four towns in Illinois: Peoria, Bloomington, Quincy, and Galesburg. They were concerned with the impact of network ties on the physicians' adoption of a new drug, tetracycline. These views are advice, discussion, and friend.

Celegans multiplex connectome network: this dataset considered different types of genetic interactions for organisms in the Biological General Repository for Interaction Datasets (BioGRID, thebiogrid.org), a public database that archives and disseminates genetic and protein interaction (ElectrJ, MonoSyn, and PolySyn) data from humans and model organisms.

These networks have been used as benchmark datasets for evaluating multiplex network analysis methods. In addition, the CKM dataset has ground-truth information about the community label of nodes. Therefore, HMNE performs performance testing of the cross-domain link prediction task on all datasets and performs performance testing of the shared community detection task on the CKM dataset.

5.1.2. Private Dataset. This dataset is a two-layer network constructed from two views, one of which is a coauthor network constructed in the form of author co-occurrence from common paper data (from AMiner). Another view is to take the author of the dissertation as the central node from each acknowledgment chapter of the dissertation data, the named entity (including tutor, teacher, classmate, or family member) identified in the acknowledgment text as the neighbor node, and the co-occurrence of the entity as the edge constructed from the center network (ego network). Based on the acknowledgment text of the dissertation and paper data, the acknowledgment layer network and coauthor layer network of the Ack-co-author dataset are constructed, respectively.

5.2. Baseline Methods. In these experiments, we test 14 other comparison algorithms: 11 baseline methods with the same parameters and dimensions and 3 traditional methods. The explanations of these baseline methods are as follows. Some of these methods can be used to test two tasks simultaneously. Other methods can only be suited for one of two tasks. The details of baseline methods are as follows:

(i) CN (common neighbor): it captures the notion that two nodes that have a common neighbor may be introduced by that neighbor. It has the effect of

Input: graph $G = \langle \mathscr{V}, \mathscr{E}, L, \mathscr{X} \rangle$; neural layer number $K \geq 3$ for GCC and GDC, graph convolution/deconvolution kernel Θ, Θ_d, iteration times T.
Output: $\mathscr{H}$: the node embeddings of multiplex network G

begin
  Initialize all parameters for GCC and GDC with K neural layers, respectively.
  $t = 1$
  while $t \geq T$ or not converge **do**
    for l in L **do**
      Sample nodes and calculate $P_{(l,\cdot)}$ in layer l based on equation (15).
      Generate convolution embedding $\mathbf{Z}^l$ using χ and G^l by equation (5)
      Readout the embedding $\mathscr{H}^l$ of layer l by equation (14)
      Generate disentangled embedding $\overline{\mathbf{Z}}^l$ using $\mathbf{Z}^l$ and G^l by equation (6).
    end
    Calculate P_{true} by equations (11) and (13).
    Calculate P_{pred} by equation (16).
    Update Θ and Θ_d by minimizing equations (7) and (17).
    Generate the reconstruction attributes $\widetilde{\chi}$ by $\overline{\mathbf{Z}}^l$ and equation (18).
    Update Θ, Θ_d, and W by minimizing equation (19).
    $t+ = 1$
  end
  Incorporate node embedding $\mathscr{H}$ by equation (9).
end
return the node representation $\mathscr{H}$.

Algorithm 1: HMNE model.

Table 2: Basic statistics about different multiplex networks used in this study.

Name	Nodes	Edges	Layers	Description
Vickers	29	740	3	Class: 316; best friend: 226; work: 198
CS-Aarhus	61	620	5	Facebook: 193; leisure: 124; work: 21; coauthor: 87; lunch: 195
London	369	441	3	Tube: 312; overground: 82; DLR: 46
CKM	246	1551	3	Advice: 480; discussion: 565; friend: 506
Celegans	279	5863	3	ElectrJ: 1031; MonoSyn: 1639; PolySyn: 3193
Ack-co-author	3383	29128	2	Acknowledgment: 1733; coauthor: 1285

"closing a triangle" in the graph and likes a common mechanism in real life.

(ii) JC (Jaccard coefficient): it is a measure used for gauging the similarity and diversity of sample sets and is defined as the size of the intersection divided by the size of the union of the sample sets.

(iii) AA (Adamic/Adar): it is a measure to predict links, according to the number of shared links between two nodes. It is defined as the sum of the inverse logarithmic degree centrality of the neighbors shared by the two nodes.

(iv) AAMT [54]: it is a link prediction method for multiplex networks based on the Adamic/Adar coefficient neighbor similarity, which considers the intensity and structural overlap of multiplex links simultaneously.

(v) Node2vec [33]: it adds a pair of parameters to achieve BFS and DFS sampling process on the single-layer network. It makes it better for capturing the role of nodes, such as hubs or tail users.

(vi) OhmNet [18]: it is a node embedding method for multiplex networks, where hierarchy information is used to model dependencies between the layers.

(vii) PMNE [17]: it has three methods of node embedding, each of which generates a common embedding of each node by merging multiple networks. We compare these three models with other baseline methods. We denote "network aggregation," "results' aggregation," and "coanalysis model" as PMNE(n), PMNE(r), and PMNE(c), respectively.

(viii) MNE [24]: it is a scalable multiplex network embedding. It contains one high-dimensional common embedding and a lower-dimensional additional embedding for each type of relations. Then, multiple relations can be learned jointly based on a unified network embedding model.

(ix) MELL [26]: it is a novel embedding method for multiplex networks, which incorporates an idea of layer vector that captures and characterizes each layer's connectivity. This method exploits the

overall structure effectively and embeds both directed and undirected multiplex networks, whether their layer structures are similar or complementary.

(x) GraphSAGE [38]: it is a graph neural network framework for inductive representation learning on graphs. GraphSAGE is used to generate low-dimensional vector representations for nodes and is especially useful for graphs that have rich node attribute information. We use an unsupervised learning version of GraphSAGE to serve as a baseline method of the link prediction task.

(xi) GATNE-T [9]: it considers the network structure and uses base embeddings and edge embeddings to capture the influential factors between different edge types. The attention mechanism is used to capture the influential factors between different edge types.

(xii) DMGI [10]: it is a simple yet effective unsupervised network embedding method for the attributed multiplex network, inspired by Deep Graph Infomax (DGI), which maximizes the mutual information between local patches of a graph and the global representation of the entire graph.

(xiii) MV-ACM [45]: it is a novel multiview adversarial completion model (MV-ACM). Each relation space is characterized in a single viewpoint, enabling them to use the topological structural information in each view.

(xiv) GenLouvain [55]: it is a modularity-based multiplex network community detection algorithm. The algorithm not only considers the modularity within the layer but also considers the modularity between layers. By maximizing the modularity metrics, the algorithm completes the community detection task. We only use this algorithm as a baseline method for the node clustering task.

In this paper, we only apply CN, JC, AA, node2Vec, and GraphSAGE to link prediction tasks at the single layer where test edges are located at. For OhmNet, we construct a hierarchy describing relationships between different layers randomly. We regard the common embedding in the MNE algorithm as the global embedding of nodes. For MELL, we add layer-level embedding as the global-level embedding and then add it to the node-level embedding of the test node. AAMT uses the multiplexity property of nodes (interlayer information) and similarity between nodes (intralayer information) to predict the probability of link. For GATNE-T and MV-ACM, we only use the homogeneous skip-gram model for node representation learning. The categorical multislice network model is selected for GenLouvain. Besides the same walk length, walk times and embedded dimensions are set as the same parameters of HMNE, and we also set other experimental baseline methods using the default parameters, such as PMNE, MELL, and DMGI.

5.3. Experimental Setup. For implementing the network feature extraction module, we use representation learning of nodes to extract the feature of each layer. In these datasets we use, if nodes in these datasets have no attributes, we use the adjacency matrix of merged multiplex networks as the attribute information of nodes in compared experiments. The definition of the matrix is the adjacency matrix of the multilayer network after the multilayer network is aggregated or flattened (that is, the union of edges for each layer). The matrix can reflect that the topology of nodes in different networks depends on the network topology. In other words, neighbor nodes (denote node attributes) are dependent on the formation of the node topology under different semantics. We set $p = 2$ and $q = 1$ as default parameters in the biased sample process of the node2vec method. We set the number of walks to 20 and walk length to 30 for OhmNet, node2vec, PMNE (n, r, c), MNE, MELL, GraphSAGE (unsupervised vision), GATNE-T, and MV-ACM. The dimension of embedding is set to 128 for all methods. For GATNE-T, DMGI, MV-ACM, and our HMNE, the optimizer of the model is Adam, the learning rate is selected from {0.0001, 0.002}, and the batch size is 50 (except for the Vickers dataset). For three heterogeneous embedding network methods, an edge is usually input into the model as a meta-path for training. All the experiments are conducted on a Linux server with sixteen logical CPUs on Intel Xeon E5 CPU and four GTX 1080Ti GPUs. Notice that, in the community detection task, we uniformly remove the community label in the node attributes for representation learning. Although our model can alleviate the oversmoothing problem of the current graph neural network algorithm, to verify this feature of our model, we show the effect of different layers of the neural network on the model performance. According to the experiment results, it is a tradeoff between the performance and complexity of the model to use a 2-layer graph neural network in both compared experiments.

5.4. Cross-Domain Link Prediction. In this section, we perform the cross-domain link prediction task on these multiplex networks. We refer to the experimental settings of the multiplex networks of literature [45]. For the cross-domain link prediction task, we remove 20% of edges of each layer in the original network and use the area under the curve (AUC) score and adjusted mutual information (AMI) score to evaluate the performance of these algorithms for predicting missing edges in each layer. We use the residual (80%) edges of each layer for training and the 20% of edges randomly selected from each layer for testing. These node pairs in edge sets of the test set are regarded as positive examples. Then, we randomly sample an equal number of node pairs from the test set, in which no edge connecting node pairs are served as negative examples. AUC is the area under the receiver operating characteristic (ROC) curve, which is equal to the probability that a classifier ranks a randomly chosen positive example higher than a randomly chosen negative one. With Pos positive examples and Neg negative examples, AUC can be calculated by

$$\mathrm{AUC} = \frac{\sum_{i \in +} \mathrm{rank}_i - (\mathrm{Pos}\,(1 + \mathrm{Pos})/2)}{\mathrm{Pos} \times \mathrm{Neg}}. \tag{21}$$

Mutual information (MI) is also used to measure the degree of agreement between the two data distributions.

Assuming that U and Y are the distribution of N sample labels, the entropy of the two distributions is

$$\begin{aligned} P(i) &= \frac{|U_i|}{N}, \\ \widetilde{P}(j) &= \frac{|Y_j|}{N}, \\ H(U) &= \sum_{i=1}^{|U|} P(i)\log(P(i)), \\ \widetilde{H}(Y) &= \sum_{j=1}^{|Y|} \widetilde{P}(j)\log(\widetilde{P}(j)), \\ \mathrm{MI}(U,Y) &= \sum_{i=1}^{|U|}\sum_{j=1}^{|Y|} P(i,j)\log\left(\frac{\widetilde{P}(i,j)}{\widetilde{P}(i)\widetilde{P}(j)}\right), \\ \mathrm{AMI} &= \frac{\mathrm{MI} - E[\mathrm{MI}]}{\max(H(U),H(V)) - E[\mathrm{MI}]}, \end{aligned} \tag{22}$$

where $E[\mathrm{MI}]$ is an expected value of mutual information. The range of AMI values is $[-1, 1]$, and its value is larger, which means that the result is more consistent with the real situation.

We calculate the similarity between nodes by CN, JC, and AA metrics in the layer where the test node pair is located. For other single-layer network embedding methods, we train a separate embedding for each relation type of the network to predict links on the corresponding edges. It means that they do not have information from other layers of multiplex networks. We aim to verify the interlayer dependence can provide complementary structure information from other layers. In terms of node embedding methods, we use the cosine function of vectors as a similarity metric. The larger the similarity scores are, the more likely there exists a link between them.

From Table 3 and Figure 3, we can know that HMNE is significantly better than other comparison algorithms. Our model shows better performance on multiplex network datasets than single-layer methods such as CN, JC, AA, node2vec, and GraphSAGE, which directly proves that fusing different structural information by preserving the interlayer dependence property can improve the accuracy of the cross-domain link prediction task. This property of the multiplex network can provide critical complementary information from other layers. We regard OhmNet, PMNE, MNE, MELL, GATNE-T, DMGI, and HMNE as comparative experimental groups. These compared algorithms are the latest multiplex network representation learning methods to learn multiplex network representation. OhmNet and PMNE are extensions of the traditional single-layer network embedding method, but there is no direct consideration of the interlayer dependence property in the final embeddings. It leads to an inevitable loss of information in the embedding process, so the complementary information of the interlayer cannot be well preserving. For MNE and MELL methods, the common (or layer) embedding is considered based on the assumption that nodes have similar local structures in different layers. In fact, this assumption is rare, and it also affects the generalization ability of the algorithm. This process of interlayer node embedding based on common embedding can lead to distortion and inaccuracy of information. GATNE-T, DMGI, and MV-ACM are specially designed to handle such a scenario that the nodes have different types and attributes in each layer, so they cannot show excellent performance in the problem we are trying to solve. Moreover, these three methods ignore the dependence property between node attributes and the topology of the layer where the node is located. For our model, HMNE simultaneously considers intralayer, interlayer, and attribute dependence properties of nodes in the node embedding process.

5.5. Shared Community Detection. Shared community detection task aims to group similar nodes so that nodes in the same group are more similar to each other than those in different groups. In other words, each node in a multiplex network has different relations/views and only belongs to a unique community. In the CKM dataset, nodes have the global community label. For this dataset, this task is usually called a shared community detection task, which is a significant mining task in multiplex network analysis. Therefore, we treat the CKM dataset as the benchmark dataset of the shared community detection task. For these methods based on node representation learning, we use K-means++ algorithm to calculate the cluster of the final embedding of nodes. In order to evaluate fairness, we set the number of communities (clusters) to 2.

5.5.1. Evaluation Metrics. Given the ground-truth community in the real-world datasets, we use normalized mutual information (NMI) to evaluate the performance of the methods:

$$\mathrm{NMI}(X\,|\,Y) = 1 - \frac{H(X\,|\,Y) + H(Y\,|\,X)}{2}, \tag{23}$$

where X and Y denote two partitions of the network and $H(X\,|\,Y)$ denotes the normalized conditional entropy of partition X with respect to Y shown in the following equation:

$$H(X\,|\,Y) = \frac{1}{|C|}\sum_{k}\frac{H(X_k\,|\,Y)}{H(X_k)}, \tag{24}$$

where $|C|$ denotes the number of communities. The larger the NMI is, the better the result is. The value of NMI takes from 0 to 1. It is equal to 1 meaning two partitions match perfectly and is equal to 0 on the contrary.

In the domain of node clustering, the chance-corrected version of this measure is adjusted Rand index (ARI). It is known to be less sensitive to the number of parts. It is possible to say that two elements of $\mathbf{Y}$, i.e., (x, x'), are paired in $\mathbf{P}$ if they belong to the same cluster. Let $\mathbf{Q}$ and $\mathbf{U}$ be two partitions of the object set $\mathbf{Y}$. A formally formulation of the adjusted Rand index is

Table 3: Cross-domain link prediction task. All the results are the averaged AUC scores.

Node type	Network type	Algorithm	Datasets					
			Celegans	CKM	CS-Aarhus	London	Vickers	Ack-co-author
Homogeneous network	Single layer	CN	0.7467	0.6517	0.8855	0.5054	0.7932	0.5104
		JC	0.7330	0.6526	0.8883	0.5054	0.7864	0.5102
		AA	0.7524	0.6523	0.8962	0.5054	0.8145	0.6968
		Node2vec	0.7847	0.8021	0.8997	0.6816	0.6667	0.5097
		GraphSAGE	0.7629	0.8521	0.7023	0.5160	0.7571	0.3991
		AAMT	0.8604	0.8239	0.9232	0.5266	0.7389	0.6968
		OhmNet	0.8427	0.8576	0.8826	0.3580	0.7841	0.8060
		PMNE(n)	0.5012	0.4773	0.5154	0.4993	0.5013	0.4981
		PMNE(r)	0.4945	0.5043	0.5205	0.4782	0.5002	0.5076
		PMNE(c)	0.5003	0.4757	0.5047	0.5043	0.4955	0.4983
	Multiple layers	MNE	0.6313	0.7902	0.8842	0.4526	0.7048	0.7093
		MELL	0.8085	0.7599	0.9014	0.4991	0.7923	0.7227
Heterogeneous network		MV-ACM	—	0.8538	0.7966	0.7630	0.7810	—
		GATNE-T	0.8142	0.8605	0.8897	0.6631	0.8165	0.8152
Homogeneous and heterogeneous network		DMGI	0.8557	0.8535	**0.9275**	0.7501	0.8028	0.8203
Homogeneous network		HMNE	**0.8730**	**0.8669**	0.9252	**0.7785**	**0.8178**	**0.8223**

$$\mathrm{ARI} = \frac{2(ad - bc)}{b^2 + c^2 + 2ad + (a + d)(c + b)}, \tag{25}$$

where a is the number of pairs $(y, y') \in \mathbf{Y}$ that are paired in $\mathbf{Q}$ and in $\mathbf{U}$; b is the number of pairs $(y, y') \in \mathbf{Y}$ that are paired in $\mathbf{Q}$ but not paired in $\mathbf{U}$; c is the number of pairs $(y, y') \in \mathbf{Y}$ that are not paired in $\mathbf{Q}$ but paired in $\mathbf{U}$; and d is the number of pairs $(y, y') \in \mathbf{Y}$ that are neither paired in $\mathbf{Q}$ nor in $\mathbf{U}$. This index has an upper bound of 1 and takes value 0 when the Rand index is equal to its expected value.

5.5.2. Result Analysis. As shown in Table 4, HMNE shows excellent performance in the shared community detection task. Among them, HMNE has obtained the largest NMI and ARI scores. In terms of other methods, MNE and MELL learn a representation of a node separately in each layer. We sum the representations in different layers of nodes as the global embedding of nodes and compare them with our model. Therefore, the performance of MNE and MELL in this task shows that this kind of join representation learning algorithm cannot well preserve the shared community information of nodes. Compared with MV-ACM, GATNE-T, and DMGI that can handle heterogeneous networks, our model can show more excellent performance in the shared community detection task. The comparison with methods GATNE-T, MV-ACM, and DMGI that can handle heterogeneous networks shows that our model also has good performance. Unlike them, HMNE takes into account the high-order proximity property of nodes. The property encourages node embeddings for an identical community is similar. It should be noted that due to the use of the iterative strategy of maximizing modularity, GenLouvain shows competitive performance. However, GenLouvain only considers the topology of the multiplex network. HMNE can capture fine-grained semantic information by preserving the dependence property between node attributes and the topology of each layer. Compared with other algorithms, it is verified in the shared community detection task that our model can preserve the global mesoscale information of the multiplex network more effectively. We further validate that our model can more fully consider multiple properties of networks. The execution time of MV-ACM is more than 24 hours, so it does not show the final results on Celegans and Ack-co-author datasets. In general, the results of cross-domain link prediction and shared community detection tasks prove the effectiveness of our model. For the cross-domain link prediction task, the graph convolution-deconvolution component of HMNE guarantees that our model can save high-order proximity information. When there is a lack of available information within the layer, the interlayer dependence component of HMNE can provide more abundant information. For the shared community detection task, the component preserving dependence between node attributes and the topology of the layer where the node is located can obtain more fine-grained semantic information related to the layer's topology by disentangling the original attribute information.

5.6. Performance Analysis. In this section, we analyze the results of parameter analysis experiments on the CKM dataset that affect the performance of the model, mainly (1) the impact of the number of convolution (deconvolution) neural network layers on the performance of our model and (2) the impact of the embedding dimension on the performance of HMNE.

5.6.1. Effect of the Neural Layers' Number. It can be seen from the illustration in Figures 4(a) and 4(b) that HMNE can avoid the smooth transition problem caused by the increase of the number of convolution layers. For AUC and AMI scores, it clearly reveals that the performance of HMNE first increases with the increase of the number of network layers

Table 4: The ARI and NMI scores' performance of our HMNE and baseline methods on the CKM multiplex network dataset.

Algorithm	ARI	NMI
OhmNet	0.7920	0.7885
PMNE(n)	0.1733	0.1574
PMNE(r)	0.0376	0.0228
PMNE(c)	0.1582	0.1679
MNE	0.1504	0.1550
MELL	0.1728	0.1805
MV-ACM	0.8942	0.7903
GATNE-T	0.8221	0.8196
DMGI	0.8507	0.8519
GenLouvain	0.9750	0.9742
HMNE	**0.9790**	**0.9771**

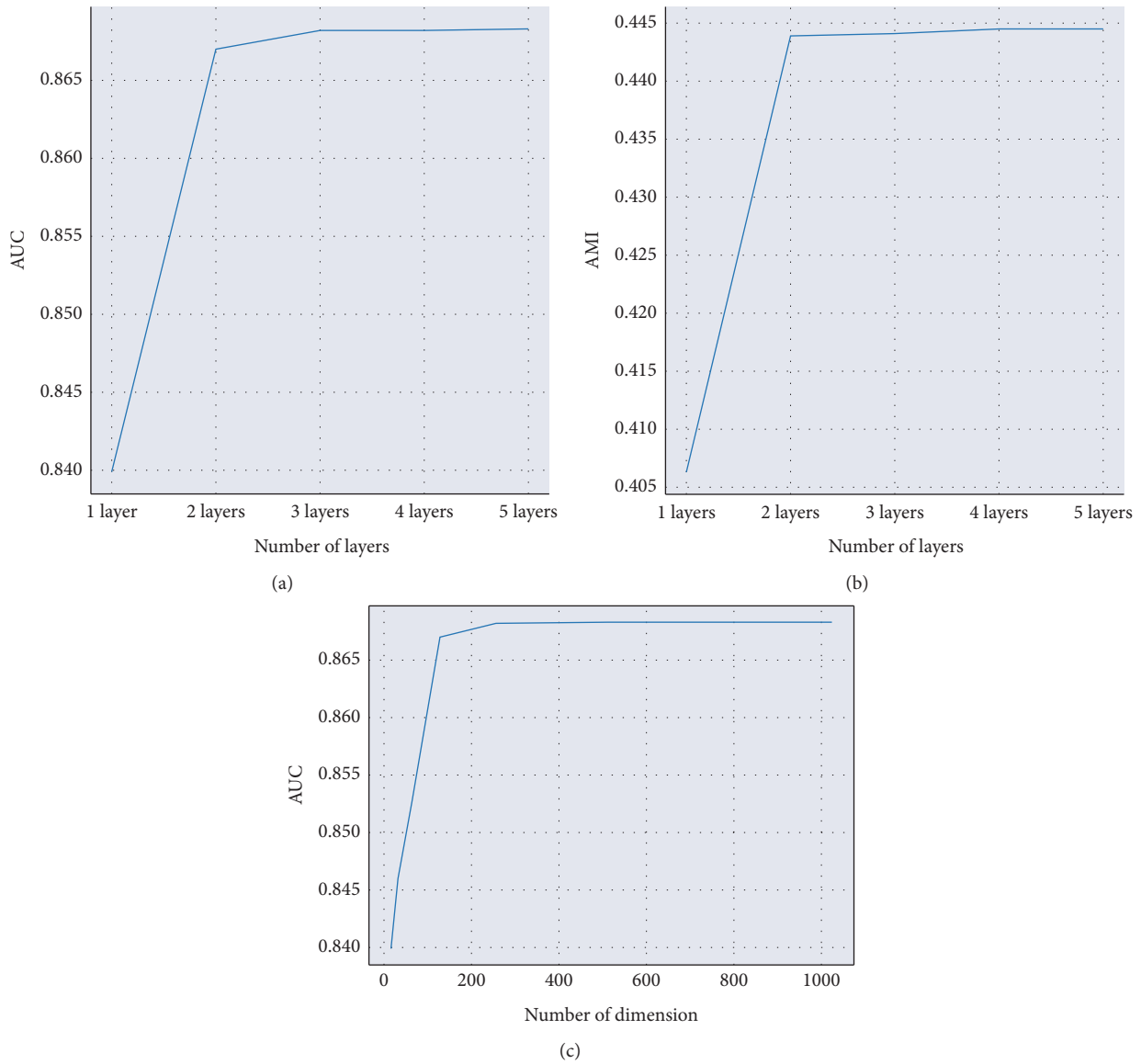


Figure 4: Impact of different experimental settings on our model's performance in the link prediction task: (a) the effect of layer's number on the AUC value, (b) the effect of layer's number on the AMI value, and (c) the effect of dimension on the AUC value.

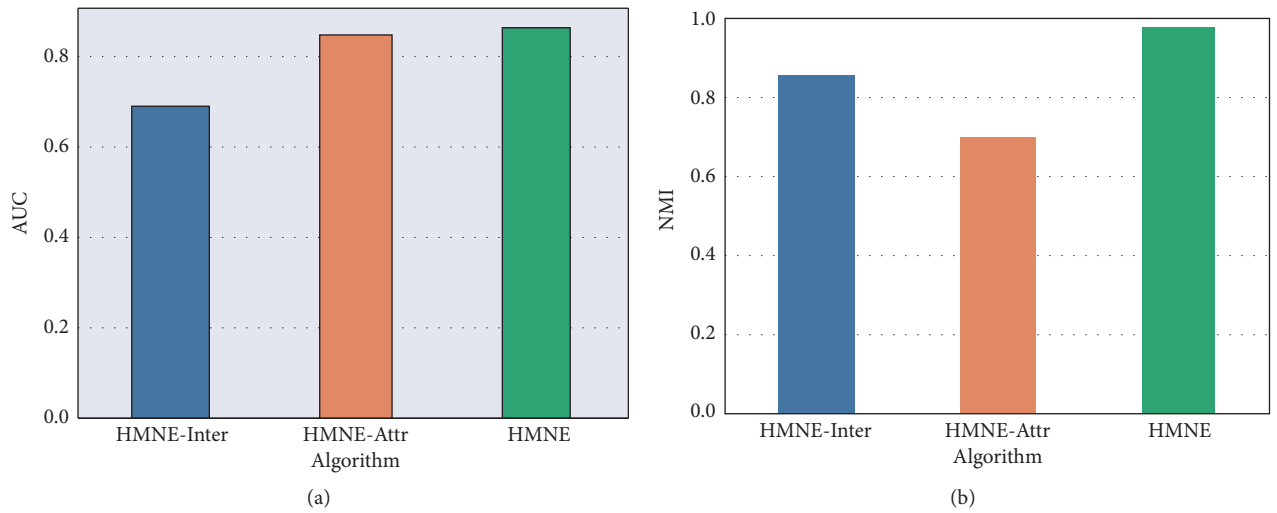


Figure 5: The effectiveness of different dependence properties in cross-domain link prediction and shared community detection tasks: (a) the effectiveness comparison in cross-domain link prediction; (b) the effectiveness comparison in the shared community detection task.

and then tends to stabilize. In other words, HMNE does not appear to be oversmoothing as the number of layers increases like other methods [56] based on graph neural networks. Therefore, our proposed HMNE can not only preserve the high-level proximity information of nodes but also avoid oversmoothing problems caused by stacking multiple neural layers.

5.6.2. Effect of the Embedding Dimension. Figure 4(c) illustrates that AUC scores of HMNE also first increase with the increase of the number of embedding dimensions and then tend to stabilize. When the embedding dimension reaches a certain level, HMNE can capture enough key information. In a certain embedding range, node embedding already contains most of the important information that is needed by some tasks. If the embedding dimension continues to increase, it will learn higher-order or more abstract information. Therefore, its performance can show a certain stable state in an interval. In this state, owing to that HMNE has similar self-supervised and autocoder structure, we believe that, with the further increase of dimensions, the objective function designed by our model will purify the original information, filter some meaningless and redundant information, and preserve fine-grained features. Therefore, as the dimension increases, the performance of the model will not show an increasing trend again in a certain dimension range.

5.7. Ablation Experiment. In this section, we will verify the effectiveness of the two properties separately by ablating the constraints of the corresponding loss function from HMNE. (1) HMNE-Inter: to verify the effect of the interlayer dependence property on HMNE, we only ablate loss function equation (17). (2) HMNE-Attr: to verify the dependence between node attributes and the topology of each layer on HMNE, we only ablate loss function equation (19). The experimental results are shown in Figure 5.

5.7.1. The Effectiveness of the Interlayer Dependence Property. As can be seen from Figure 5(a), the interlayer dependence property is critical for link prediction tasks. After removing loss function equation (17) (called HMNE-Inter), the performance of HMNE in the cross-domain link prediction task decreases more significantly than the decrease in the community detection task. The reason is that the structure information of other layers provides effective complementary information for the node pair prediction of the target layer.

5.7.2. The Effectiveness of Dependence between Node Attributes and the Topology of Each Layer. After removing loss function equation (19) (called HMNE-Attr), Figure 5(b) illustrates that HMNE-Attr decreases significantly in the community detection task. In the shared community detection task, we believe the performance of HMNE is more dependent on the attribute information of the node. However, in the link prediction task, the information provided by the dependence between node attributes and the topology of each layer is limited.

6. Conclusion

In this paper, we propose an unsupervised node embedding model for multiplex networks, called HMNE. HMNE first addresses the problem of preserving of high-order proximity information of nodes through the symmetric graph convolution-deconvolution component (SGCD). SGCD utilizes the designed graph deconvolution component (GDC) to reconstruct the input of the graph convolution component (GCC) with multiple graph convolution neural layers. It can effectively avoid the oversmoothing problem. Secondly, HMNE preserves the interlayer dependence property with interlayer complementary information of multiplex networks by our designed interlayer dependence component. When there is a lack of available information within the layer, the interlayer dependence component of HMNE can provide more abundant information from other layers (e.g., cross-domain link prediction

scenario). Finally, HMNE preserves the dependence between the node attributes and the topology of each layer through disentangled representation of attributes of nodes. It enables HMNE to have more fine-grained attributes with different semantic information of nodes associated with each layer structure. The final representation of nodes with fine-grained attribute information can perform better in downstream tasks (e.g., shared community detection scenario). Systematical experiments on six real-world networks show the excellent performance of HMNE on two downstream tasks compared with the state-of-the-art baselines. Experiments on large-scale network data based on HMNE will be our future research focus.

Data Availability

These compared methods and our code required for replicating reported results are available at https://github.com/Brian-ning/HMNE/. The public datasets can also be downloaded from https://comunelab.fbk.eu/data.php.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2018YFC0831500), the National Natural Science Foundation of China (Grant no. 61972047), and the NSFC-General Technology Basic Research Joint Funds (Grant no. U1936220).

References

[1] X. Zhu, J. Ma, X. Su et al., "Information spreading on weighted multiplex social network," *Complexity*, vol. 2019, Article ID 5920187, 15 pages, 2019.

[2] L. Gao, H. Yang, J. Wu, C. Zhou, W. Lu, and Y. Hu, "Recommendation with multi-source heterogeneous information," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence IJCAI-18*, pp. 3378–3384, Stockholm, Sweden, July 2018.

[3] A. Ghavasieh and M. De Domenico, "Enhancing transport properties in interconnected systems without altering their structure," *Physical Review Research*, vol. 2, no. 1, 2020.

[4] S. Choobdar, M. E. Ahsen, M. E. Ahsen et al., "Assessment of network module identification across complex diseases," *Nature Methods*, vol. 16, no. 9, pp. 843–852, 2019.

[5] J. Zhang and S. Y. Philip, *Broad Learning through Fusions: Broad Learning Introduction*, Springer, Berlin, Germany, 2019.

[6] M. E. J. Newman, "Network structure from rich but noisy data," *Nature Physics*, vol. 14, no. 6, pp. 542–545, 2018.

[7] Y. Zhang, J. Wu, C. Zhou, Z. Cai, J. Yang, and P. S. Yu, "Multi-view fusion with extreme learning machine for clustering," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 5, pp. 1–23, 2019.

[8] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.

[9] Y. Cen, X. Zou, J. Zhang, H. Yang, J. Zhou, and J. Tang, "Representation learning for attributed multiplex heterogeneous network," in *Proceedings of the 25th ACM SIGKDD International Conference*, pp. 1358–1368, ACM, Anchorage, AK, USA, August 2019.

[10] C. Park, D. Kim, J. Han, and H. Yu, "Unsupervised attributed multiplex network embedding," in *Proceedings of the Thirty-Fourth AAAI Conference on Artificial intelligence (AAAI)*, pp. 1–8, New York, NY, USA, February 2020.

[11] V. Gligorijevic, Y. Panagakis, and S. Zafeiriou, "Non-negative matrix factorizations for multiplex network analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 4, pp. 928–940, 2019.

[12] Y. Shi, F. Han, X. He, C. Yang, J. Luo, and J. Han, "mvn2vec: preservation and collaboration in multi-view network embedding," *Knowledge-Based Systems*, 2020.

[13] C. Park, J. Han, and H. Yu, "Deep multiplex graph infomax: attentive multiplex network embedding using global information," *Knowledge-Based Systems*, vol. 197, Article ID 105861, 2020.

[14] Y. Ouyang, B. Guo, X. Tang, X. He, J. Xiong, and Z. Yu, "Learning cross-domain representation with multi-graph neural network," *Machine Learning*, 2019.

[15] L. Yu, S. Pei, C. Zhang et al., "Self-supervised smoothing graph neural networks," 2020, http://arxiv.org/abs/2009.00934.

[16] F. Battiston, V. Nicosia, and V. Latora, "The new challenges of multiplex networks: measures and models," *The European Physical Journal Special Topics*, vol. 226, no. 3, pp. 401–416, 2017.

[17] W. Liu, P.-Y. Chen, S. Yeung, T. Suzumura, and L. Chen, "Principled multilayer network embedding," in *Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 134–141, IEEE, New Orleans, LA, USA, November 2017.

[18] M. Zitnik and J. Leskovec, "Predicting multicellular function through multi-layer tissue networks," *Bioinformatics*, vol. 33, no. 14, pp. i190–i198, 2017.

[19] P. Zhou, L. Du, X. Li, Y.-D. Shen, and Y. Qian, "Unsupervised feature selection with adaptive multiple graph learning," *Pattern Recognition*, vol. 105, pp. 107375–107389, 2020.

[20] M. S. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *Proceedings of the Semantic Web—15th International Conference (ESWC) 2018*, vol. 10843, pp. 593–607, Springer, Heraklion, Greece, June 2018.

[21] S. Vashishth, S. Sanyal, V. Nitin, and P. Talukdar, "Composition-based multi-relational graph convolutional networks," in *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, Addis Ababa, Ethiopia, April 2020.

[22] D. Yu, Y. Yang, R. Zhang, and Y. Wu, "Generalized multi-relational graph convolution network," 2020, http://arxiv.org/abs/2006.07331.

[23] J. Li, C. Chen, H. Tong, and H. Liu, "Multi-layered network embedding," in *Proceedings of the 2018 SIAM International Conference on Data Mining*, pp. 684–692, SIAM, San Diego, CA, USA, May 2018.

[24] H. Zhang, L. Qiu, L. Yi, and Y. Song, "Scalable multiplex network embedding," in *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, vol. 18, pp. 3082–3088, Stockholm, Sweden, July 2018.

[25] Y. Ma, Z. Ren, Z. Jiang, J. Tang, and D. Yin, "Multi-dimensional network embedding with hierarchical structure," in *Proceedings of the Eleventh ACM International Conference*

on Web Search and Data Mining, pp. 387–395, ACM, Los Angeles, CA, USA, February 2018.

[26] R. Matsuno and T. Murata, "Mell: effective embedding method for multiplex networks," in *Proceedings of the Companion the Web Conference (WWW)*, pp. 1261–1268, International World Wide Web Conferences Steering Committee, Lyon, France, April 2018.

[27] A. Mahmoudi, M. R. Yaakub, and A. A. Bakar, "The relationship between online social network ties and user attributes," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 3, pp. 15–26, 2021.

[28] S. Ali, M. Shakeel, I. Khan, S. Faizullah, and M. Khan, "Predicting attributes of nodes using network structure," *ACM Transactions on Intelligent Systems and Technology*, vol. 12, no. 2, pp. 1–23, 2020.

[29] J. Li, R. Guo, C. Liu, and H. Liu, "Adaptive unsupervised feature selection on attributed networks," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 92–100, Anchorage, AK, USA, August 2019.

[30] P. Velickovic, W. Fedus, W. L. Hamilton, P. Lio, Y. Bengio, and R. D. Hjelm, "Deep graph infomax," in *Proceedings of the International Conference on Learning Representations*, New Orleans, LA, USA, May 2019.

[31] C. Shi, Y. Li, J. Zhang, Y. Sun, and S. Y. Philip, "A survey of heterogeneous information network analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, 2016.

[32] P. Bryan, R. Al-Rfou, and S. Skiena, "Deepwalk: online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 701–710, ACM, New York, NY, USA, August 2014.

[33] A. Grover and J. Leskovec, "node2vec: scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864, ACM, San Francisco, CA, USA, August 2016.

[34] W. Gu, L. Gong, X. Lou, and J. Zhang, "The hidden flow structure and metric space of network embedding algorithms based on random walks," *Scientific Reports*, vol. 7, no. 1, pp. 13114–13125, 2017.

[35] Y. Dong, N. V. Chawla, A. Swami, Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: scalable representation learning for heterogeneous networks," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 135–144, Halifax, Canada, August 2017.

[36] L. F. R. Ribeiro, P. H. P. Saverese, and D. R. Figueiredo, "struc2vec: learning node representations from structural identity," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 385–394, ACM, Halifax, Canada, August 2017.

[37] T. N. Kipf and M. Welling, "Variational graph auto-encoders," in *Proceedings of the NeurIPS Workshop on Bayesian Deep Learning (NeurIPS-16 BDL)*, Barcelona, Spain, December 2016.

[38] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, pp. 1024–1034, IEEE, New York, NY, USA, 2017.

[39] X. Luo and H. H. Zhuo, "Bigsage: unsupervised inductive representation learning of graph via bi-attended sampling and global-biased aggregating," 2019.

[40] Y. Xiao, D. Chen, S. Wei, Q. Li, H. Wang, and M. Xu, "Rumor propagation dynamic model based on evolutionary game and anti-rumor," *Nonlinear Dynamics*, vol. 95, no. 1, pp. 523–539, 2019.

[41] Y. Xiao, W. Li, S. Qiang, Q. Li, H. Xiao, and Y. Liu, "A rumor & anti-rumor propagation model based on data enhancement and evolutionary game," *IEEE Transactions on Emerging Topics in Computing*, p. 1, 2020.

[42] Y. Xiao, Q. Yang, C. Sang, and Y. Liu, "Rumor diffusion model based on representation learning and anti-rumor," *IEEE Transactions on Network and Service Management*, vol. 17, no. 3, pp. 1910–1923, 2020.

[43] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: a survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[44] N. Ning, F. Long, C. Wang, Y. Zhang, Y. Yang, and B. Wu, "Nonlinear structural fusion for multiplex network," *Complexity*, vol. 2020, Article ID 7041564, 17 pages, 2020.

[45] K. Zhao, T. Bai, B. Wu et al., "Deep adversarial completion for sparse heterogeneous information network embedding," in *Proceedings of the Web Conference 2020*, pp. 508–518, Taipei, Taiwan, April 2020.

[46] W. Yuan, K. He, C. Shi et al., "Multi-view network embedding with node similarity ensemble," *World Wide Web*, vol. 23, pp. 2699–2714, 2020.

[47] M. Qu, J. Tang, J. Shang, X. Ren, M. Zhang, and J. Han, "An attention-based collaboration framework for multi-view network representation learning," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1767–1776, Singapore, November 2017.

[48] A. Zhiyuli, X. Liang, and Y. Chen, "Hsem: highly scalable node embedding for link prediction in very large-scale social networks," *World Wide Web*, vol. 22, pp. 2799–2824, 2019.

[49] Y. Sun, S. Wang, T.-Yu Hsieh, X. Tang, and V. Honavar, "Megan: a generative adversarial network for multi-view network embedding," 2019, http://arxiv.org/abs/1909.01084.

[50] H. Wei, Z. Pan, G. Hu et al., "Attributed network representation learning via deepwalk," *Intelligent Data Analysis*, vol. 23, no. 4, pp. 877–893, 2019.

[51] X. Chu, X. Fan, D. Yao, Z. Zhu, J. Huang, and J. Bi, "Cross-network embedding for multi-network alignment," in *Proceedings of the World Wide Web Conference*, pp. 273–284, San Francisco, CA, USA, May 2019.

[52] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017, http://arxiv.org/abs/1609.02907.

[53] D. Hjelm, A. Fedorov, S. Lavoie-Marchildon et al., "Learning deep representations by mutual information estimation and maximization," in *Proceedings of the International Conference for Learning Representations 2019*, New Orleans, LA, USA, May 2019.

[54] D. Hristova, A. Noulas, C. Brown, M. Musolesi, and C. Mascolo, "A multilayer approach to multiplexity and link prediction in online geo-social networks," *EPJ Data Science*, vol. 5, no. 24, pp. 1–17, 2016.

[55] L. G. S. Jeub, M. Bazzi, I. S. Jutla, and P. J. Mucha, "A generalized Louvain method for community detection implemented," 2011–2019, https://github.com/GenLouvain/GenLouvain.

[56] C. Cai and Y. Wang, "A note on over-smoothing for graph neural networks," 2020, http://arxiv.org/abs/2006.13318.

Hindawi
Complexity
Volume 2021, Article ID 8834652, 14 pages
https://doi.org/10.1155/2021/8834652

Research Article

No-Reference Stereoscopic Image Quality Assessment Based on Binocular Statistical Features and Machine Learning

Peng Xu,[1] Man Guo,[2] Lei Chen,[1] Weifeng Hu,[1] Qingshan Chen,[1] and Yujun Li[1]

[1]*School of Information Science and Engineering, Shandong University, Qingdao 266237, China*
[2]*Torch High Technology Industry Development Center, Ministry of Science & Technology, Beijing 100045, China*

Correspondence should be addressed to Lei Chen; leiguanglg@hotmail.com and Yujun Li; liyujun@sdu.edu.cn

Received 26 September 2020; Revised 16 December 2020; Accepted 30 December 2020; Published 29 January 2021

Academic Editor: Jia Wu

Learning a deep structure representation for complex information networks is a vital research area, and assessing the quality of stereoscopic images or videos is challenging due to complex 3D quality factors. In this paper, we explore how to extract effective features to enhance the prediction accuracy of perceptual quality assessment. Inspired by the structure representation of the human visual system and the machine learning technique, we propose a no-reference quality assessment scheme for stereoscopic images. More specifically, the statistical features of the gradient magnitude and Laplacian of Gaussian responses are extracted to form binocular quality-predictive features. After feature extraction, these features of distorted stereoscopic image and its human perceptual score are used to construct a statistical regression model with the machine learning technique. Experimental results on the benchmark databases show that the proposed model generates image quality prediction well correlated with the human visual perception and delivers highly competitive performance with the typical and representative methods. The proposed scheme can be further applied to the real-world applications on video broadcasting and 3D multimedia industry.

1. Introduction

During the past few decades, there has been an exponential increase of stereoscopic images and videos in 3D display market [1]. However, due to various 3D quality factors [2, 3] including binocular rivalry, visual comfort, and depth perception, the visual quality assessment of stereoscopic images is much more complex and relatively less researched than the traditional 2D image quality evaluation. To address these challenges, we require a deeper understanding of binocular vision mechanisms and interactions for the quality prediction of distorted stereoscopic images.

There are mainly two groups of methods on 3D image quality assessment (IQA): subjective quality evaluation by human observer [4] and objective quality evaluation by devised metric used to simulate human perceptual judgements [5]. Since the human eyes are the final receiver of visual information, the subjective evaluation can directly reflect the human visual perception and is accurate and effective to evaluate the visual quality. However, the subjective evaluation involves many participants in the course of experiments, which is time-consuming and costly. Therefore, it is unrealistic to implement it in many scenarios like real-time evaluation [6]. As a result, it is in urgent demand to propose objective methods that can effectively evaluate the human perceptual quality of stereoscopic images.

Based on the volume of accessible information in the images, existing objective quality assessment metrics can be generally divided into three categories: full-reference (FR) [7, 8], reduced-reference (RR) [9], and no-reference/blind (NR) methods [10, 11]. When the reference contents are accessible, the FR method can offer more accurate quality assessment. Early approaches for FR 3D-IQA directly stemmed from 2D quality metrics [12]. Conventionally, a straightforward way is to apply the 2D-IQA metrics to both views of a 3D image independently and then integrate the two 2D quality scores into a final 3D quality score. Several 3D-IQA methods [13, 14] were proposed by introducing the associated disparity or depth map into the 3D image quality

model. These research findings indicated that a satisfactory result can be obtained if the disparity images and reference images are combined appropriately. Afterwards, more sophisticated algorithms were developed based on the binocular vision properties. For example, Lin and Wu [15] revisited the physiological discoveries of binocular vision and incorporated the binocular integration into the existing 2D quality metrics for measuring the quality of stereoscopic images. Shao et al. [16] classified the stereoscopic images into noncorresponding, binocular fusion, and binocular suppression regions. Each region was evaluated individually according to its binocular perception property. In our previous work [17], we proposed a full-reference quality evaluator by considering the local and global qualities of 3D images. The experimental results showed its good performance in terms of stereoscopic image quality assessment.

Since pristine reference images are rarely available in practical applications [18, 19], the NR algorithms are potentially much more feasible solutions. They can give quality evaluation without any information extracted from the corresponding pristine image. The NR 3D-IQA is still preliminary, and a limited number of blind 3D-IQA algorithms have been developed. Inspired by the human visual system, Chen et al. [20] proposed a no-reference binocular image quality assessment method for natural stereopairs. The proposed method extracted both 2D and 3D natural statistical features from a stereopair and utilized these statistical features and the binocular rivalry for 3D image quality prediction. Ryu and Sohn [21] investigated the relationship between visual information and binocular quality perception and developed an NR quality evaluation algorithm for 3D images. The scores of perceptual blockiness and blurriness were combined into an overall quality index based on the binocular perception models. Shen et al. [22] devised a no-reference quality scheme for stereoscopic images based on the visual perceptual characteristics. Three types of features relating to image distortion, depth perception, and binocular disparity were used to map the human opinion scores. Other relevant works can be found in references [23–25].

Recently, machine learning techniques have achieved great success and been widely applied to various research fields [26–28]. One of the advantages of applying machine learning to quality evaluation is that it can directly take original image data as input and then combine feature learning with quality regression in the training procedure [29, 30]. Kang et al. [31] applied convolution neural network (CNN) to image quality assessment. They devised a shallow network which extracts quality-predictive features from image patches. Several NR algorithms for 3D-IQA using deep learning have been developed. Oh et al. [32] reported a novel deep learning method for NR 3D-IQA in terms of local to global feature aggregation. Zhou et al. [33] proposed a dual-stream interactive network for stereoscopic image quality assessment. In our previous work [34], we developed a no-reference quality prediction scheme for 3D images based on binocular features and support vector regression (SVR). The scheme showed its effectiveness, but the performance in terms of prediction accuracy and time complexity needs to be further improved. More efficient methods for stereoscopic image quality assessment should be explored to address these limitations.

It is challenging for the NR algorithms to have the assessment accuracy as good as can be obtained with the FR quality evaluation methods. Moreover, 3D image quality databases generally lack large-scale training images with subjective quality scores, which limit the performance of these algorithms using deep neural networks. Other techniques should be explored and worthy of further research. We are motivated to tackle these limitations for 3D image quality assessment. In this paper, inspired by the research findings on the human binocular visual system, we try to simulate the perceptual mechanism of binocular vision. We primarily work on extracting certain types of binocular features from distorted stereoscopic image and constructing a statistical regression model to map these quality-aware features to the human perceptual judgements. The main contributions of this work are as follows:

(1) Different from other related studies [33, 35], the novelty of our work lies in that we propose to adopt the effective binocular statistical features from the fusion and difference maps of a stereopair for stereoscopic image quality prediction.

(2) We have demonstrated that appropriate combination of binocular features and binocular energy can greatly promote the performance of 3D image quality evaluation.

(3) Compared with other typical and representative methods, the proposed scheme achieves higher consistent alignment with human subjective assessment and has lower time complexity. The experimental results show that our scheme can accurately estimate the perceptual quality of distorted stereoscopic images and has promising generalization ability.

The remainder of this paper is organized as follows. Section 2 introduces some fundamental knowledge about binocular visual perception. Section 3 presents the proposed quality assessment scheme for stereoscopic images in detail. Section 4 gives the experimental results and performance analysis of the proposed scheme and the comparison with other related algorithms. Finally, Section 5 concludes the paper with possible ideas for future work.

2. Foundation for Binocular Visual Perception

It has been known that the binocular vision is a complex visual process that requires the brain and both eyes working together to produce clear vision. Figure 1 describes a simplified framework of two important visual neural pathways for the binocular visual system. The ventral stream starts from the primary visual cortex V1 and goes through V2 and V3 to V4 area. The functions of ventral stream are about the recognition and perception behaviors. The dorsal stream begins from the V1 area, goes through V2 and V3 to V5 area. The visual information-guided interactions occur in dorsal

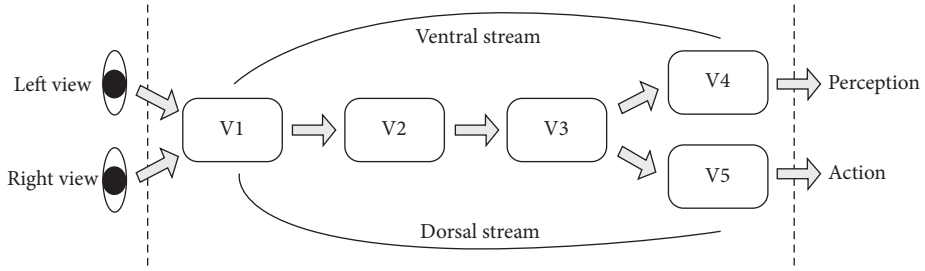


Figure 1: A simplified illustration of visual cortex with ventral stream and dorsal stream.

stream [15]. Regarding the detailed functions of each visual area, please refer to the binocular visual perception book [36] for further information.

The visual cortex plays an important role in our binocular visual perception, and it has been demonstrated that the primary visual cortex (V1) is mainly responsible for the human visual system (HVS) [37]. In the V1, simple and complex receptive fields are usually characterized to understand the behavior of visual perception. According to visual psychophysical study, two visual phenomena usually occur in the process of binocular visual response: binocular rivalry and binocular fusion. When the two eyes view mismatched images at the same retinal location, one experiences binocular rivalry. As a result of competition between the eyes, binocular rivalry involves reciprocal inhibition between the monocular channels. When two slightly different retinal signals can be perceived by two eyes, one experiences binocular fusion. During fusion, two retinal points are integrated into one single perception, superimposing and combining similar contents from the two views. Therefore, the binocular vision can be generally considered as a combination of binocular rivalry and binocular fusion.

As a significant content of primitives in V1, image structural information is closely related to image visual quality. And the degradation of perceptual quality can be reflected via the change of image structural information. Previous studies [29, 38] highlighted the significance of image structural information for image quality assessment. The gradient magnitude (GM) and Laplacian of Gaussian (LOG) are basic elements that are commonly used to represent image semantic structures [39]. More importantly, during 3D visual stimuli processing, binocular fusion and disparity responses are primitively formed in the V1 cortical area. The visual signals from the binocular summation and subtraction channels are multiplexed, and then each neuron in V1 receives a weighted sum of the visual stimuli from these two channels [40]. Motivated by these research results, in this paper, we extract the GM and LOG features from a stereopair and its fusion and difference maps as binocular features. In the following section, we will describe our proposed quality prediction model for distorted stereoscopic images in detail.

3. The Proposed No-Reference Quality Assessment Scheme

Figure 2 illustrates the architecture of the proposed scheme for stereoscopic image quality prediction. Given an original stereopair, we first generate the fusion map and difference map and extract the binocular statistical features from them as basic feature vectors. Then, we calculate the binocular energy responses from the local amplitude and local phase of the stereopair as quality-aware features. Finally, we employ an extreme learning machine method to map these features of distorted stereopair to its human perceptual quality score.

3.1. Binocular Feature Extraction. As can be seen from Figure 3, the fusion maps and difference maps of the left- and right-view images with different distortion types are discriminative, which can be utilized for extracting effective quality features. Specifically, the fusion map reflects the fusion ability of the left and right stereo-halves, while the difference map reveals the disparity information of a stereopair.

As discussed in Section 2, the gradient magnitude (GM) and Laplacian of Gaussian (LOG) features can be adopted to build the basic elements of image semantic structures, and they are hence closely related to the perceptual quality of natural images. The Gaussian derivative functions can model the receptive field responses of neurons along the visual pathway [41]. Therefore, we compute the GM and LOG maps using the first- and second-order derivatives of a circularly symmetric 2D Gaussian function G defined as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \tag{1}$$

where x and y represent the horizontal and vertical directions, respectively. The parameter σ is the standard deviation. Then, we calculate the first-order partial derivative of $G(x, y, \sigma)$ with respect to x or y by

$$\frac{\partial G(x, y, \sigma)}{\partial d} = -\frac{1}{2\pi\sigma^2} \frac{d}{\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \tag{2}$$

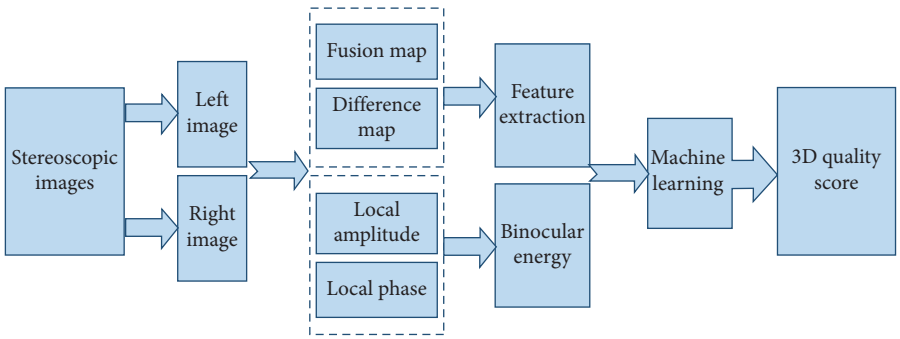


Figure 2: The architecture of our proposed quality assessment scheme for stereoscopic images.

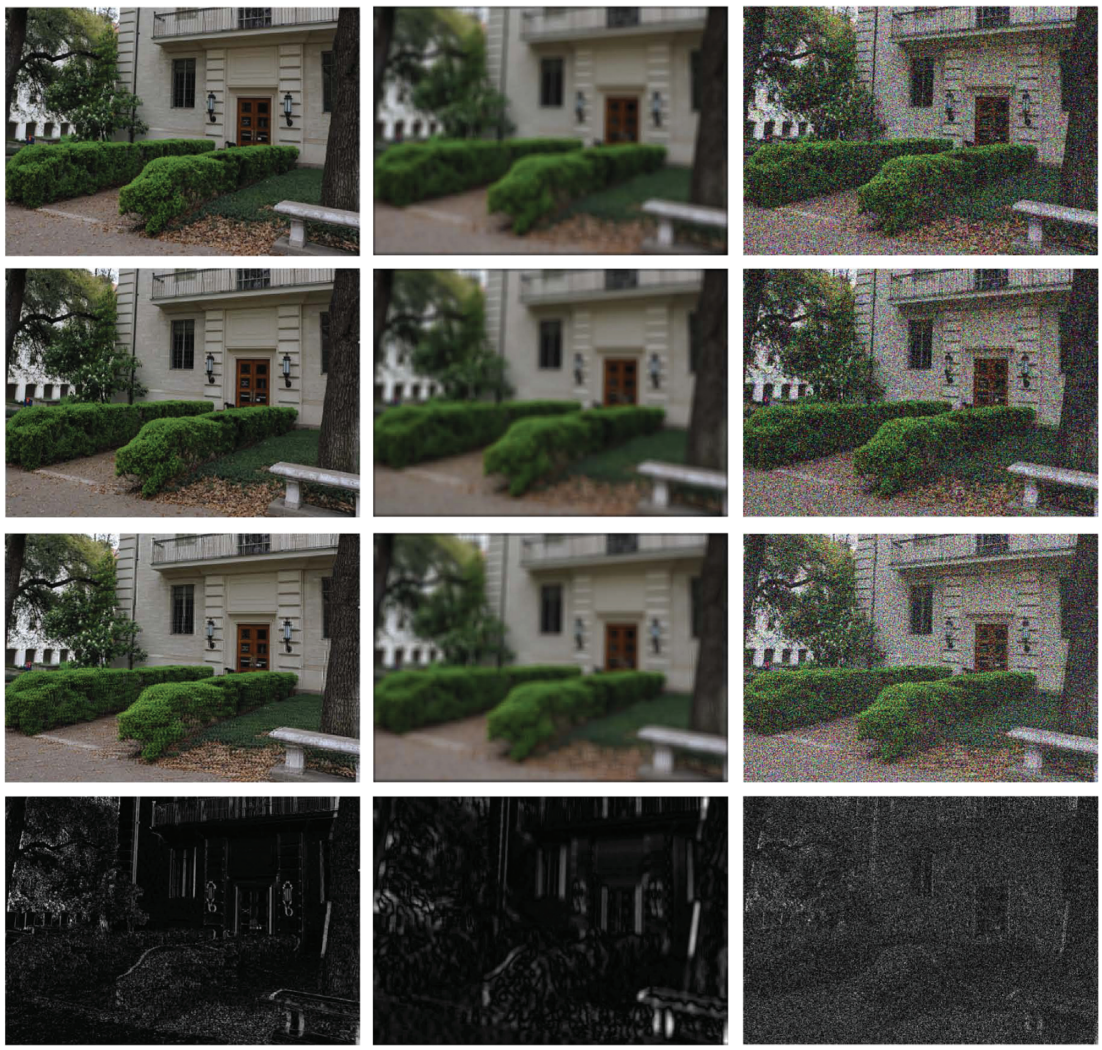


Figure 3: Examples of distorted stereoscopic images with different distortion types and the corresponding fusion and difference maps: the original stereoscopic image (left), distorted stereoscopic image with Gaussian blur (middle), and distorted stereoscopic image with white Gaussian noise (right). The first to the last rows show the left-view images, right-view images, fusion maps, and difference maps, respectively.

where $d \in \{x, y\}$ is the Gaussian partial derivative filter applied along the horizontal x or vertical y direction. An image is denoted by I; thus, the GM map of the image I can be obtained by

$$\mathrm{GM}_v = \sqrt{\left(I_v \otimes \frac{\partial G}{\partial x}\right)^2 + \left(I_v \otimes \frac{\partial G}{\partial y}\right)^2}, \tag{3}$$

where the symbol $\otimes$ represents the convolution operation. $v \in \{l, r\}$, where l and r refer to the left and right views of a stereopair, respectively. Similarly, the LOG filter, corresponding to the second-order Gaussian partial derivative, is defined as follows:

$$\begin{aligned} h_{\mathrm{LOG}}(x, y, \sigma) &= \frac{\partial^2 G(x, y, \sigma)}{\partial^2 x} + \frac{\partial^2 G(x, y, \sigma)}{\partial^2 y} \\ &= \frac{1}{2\pi\sigma^2} \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \end{aligned} \tag{4}$$

Accordingly, we estimate the LOG map of the left and right views by

$$\mathrm{LOG}_v = I_v \otimes h_{\mathrm{LOG}}. \tag{5}$$

Subsequently, a joint adaptive normalization procedure [39] is employed to normalize the GM and LOG coefficients for stable statistical image representations. Previous works [42, 43] have revealed that the overall quality of a distorted stereopair cannot be accurately calculated by directly averaging the qualities of the left- and right-view images, especially for an asymmetrically distorted 3D image. In practical application, to simulate the binocular rivalry (BR) phenomenon, the basis of weighting factors for the quality-predictive feature vectors of a stereopair can be defined as follows:

$$\begin{aligned} w_l &= \frac{(e_l)^\alpha}{\left[(e_l)^\alpha + (e_r)^\alpha\right]}, \\ w_r &= \frac{(e_r)^\alpha}{\left[(e_l)^\alpha + (e_r)^\alpha\right]}, \end{aligned} \tag{6}$$

where w_l and w_r denote the weights for the distorted left and right views, which can reflect the binocular contrast combination to a certain extent. e_l and e_r represent the local energy variances of the left and right images for a stereopair, respectively. The intensity adjusting parameter α is empirically set to 3 in the experiment. Therefore, the basic feature vectors of the gradient magnitude and Laplacian of Gaussian responses for a stereoscopic image can be calculated by

$$\begin{aligned} S_{\mathrm{GM}} &= \sum_{x,y} w_l \mathrm{GM}_l(x, y) + \sum_{x,y} w_r \mathrm{GM}_r(x, y), \\ S_{\mathrm{LOG}} &= \sum_{x,y} w_l \mathrm{LOG}_l(x, y) + \sum_{x,y} w_r \mathrm{LOG}_r(x, y), \end{aligned} \tag{7}$$

where $\mathrm{GM}_{l/r}(x, y)$ and $\mathrm{LOG}_{l/r}(x, y)$ denote the gradient magnitude and Laplacian of Gaussian for the left and right image, respectively. The features of GM and LOG responses are utilized to represent the visual semantic structures of the first-order and second-order binocular combination.

Finally, combined with the GM and LOG features of the fusion and difference maps, the binocular feature vectors used for further data training can be expressed by

$$V = \left[S_{\mathrm{GM}}, S_{\mathrm{LOG}}, F_{\mathrm{GM}}, F_{\mathrm{LOG}}, D_{\mathrm{GM}}, D_{\mathrm{LOG}}\right], \tag{8}$$

where $F_{\mathrm{GM/LOG}}$ and $D_{\mathrm{GM/LOG}}$ are the GM/LOG features of the fusion and difference maps, respectively.

3.2. Binocular Energy Response. The above extracted features are mainly utilized to indicate the visual sensitivity of binocular rivalry. Neurological research has reported that the human binocular vision phenomenon is a complicated process with combination of binocular rivalry, binocular fusion, and other factors [44]. The binocular fusion also contributes significantly to human visual perception besides the binocular rivalry. Previous research findings [45] indicated that the binocular energy responses play critical roles in representing binocular visual perception, especially for binocular fusion. In this paper, the binocular energy responses are obtained from the local magnitude and local phase of a stereopair.

In the proposed scheme, the left and right images of a stereopair are first processed using the log-Gabor filter. Here, we define $\alpha_{s,o}$ to represent the responses on different scales, where s is the spatial scale index. And we let $\beta_{s,o}$ denote the responses along different orientations, where o is the orientation scale index. The detailed description of this log-Gabor filter can be referred to the work in [46]. According to the given scale and orientation, the local amplitude at location x on scale s and along orientation o can be defined as

$$\mathrm{LA}_{s,o}(x) = \sqrt{\alpha_{s,o}^2(x) + \beta_{s,o}^2(x)}. \tag{9}$$

With the sum of the local amplitudes on all the scales along the orientation o_m [46], the local amplitude can be calculated by

$$\mathrm{LA}(x) = \sum_s \mathrm{LA}_{s,o_m}(x), \tag{10}$$

where o_m is a parameter used to indicate the orientation with the maximum phase congruency value. Similar to the local amplitude, the local phase can be obtained by the angle along the orientation [46]:

$$\mathrm{LP}(x) = \arctan\left(\frac{\sum_s \beta_{s,o_m}(x)}{\sum_s \alpha_{s,o_m}(x)}\right). \tag{11}$$

Based on previous works on binocular vision energy [16, 34], the left-view response and right-view response of a stereopair can be defined as follows:

$$\begin{aligned} C_l(x) &= \mathrm{LA}_l(x) \cdot \exp\left(\mathrm{LP}_l(x)\right), \\ C_r(x) &= \mathrm{LA}_r(x) \cdot \exp\left(\mathrm{LP}_r(x)\right). \end{aligned} \tag{12}$$

The right-view response $C_r(x)$ usually can be taken as a shifted transformation of the left-view response $C_l(x)$. The disparity d is defined as the difference between the locations of associated points in the left- and right-view responses. By considering a simple binocular cell with the left and right receptive fields, the binocular energy response E for a stereoscopic image pair can be calculated by

$$E = \sum_{x} \left\| C_l(x) + C_r(x+d) \right\|^2. \tag{13}$$

Finally, combined with the binocular statistical features, the overall quality-predictive features are $F = [V, E]$, which are fed into the following quality prediction for model learning. The weights among them are determined in the learning process.

3.3. Quality Prediction Model Learning. A number of training methods can be utilized to map the quality-predictive features of a stereopair to its corresponding subjective quality score, such as support vector regression (SVR) [47] and neural networks (NNs) [48]. SVR requires complex training algorithms and involves a quadratic programming problem. Neural networks have the difficulties of local minima, learning epochs, and slow convergence. An important question is that neural networks or training-based methods usually need large quantities of labeled training samples, while 3D image quality databases generally lack large-scale training images with subjective quality scores, which limits the performance of the methods using deep neural networks. In recent years, the extreme learning machine (ELM) [49] has attracted considerable attention and has been demonstrated as an effective and efficient technique in many applications, such as pattern recognition [50] and quality evaluation [51]. The ELM has advantages of faster learning speed, higher learning accuracy, and improved generalization. The weights between the input and hidden layers can be selected randomly and independent of the training data, and layer-by-layer back propagated tuning is not required [35]. Motivated by these unique properties, we try to employ the ELM for feature mapping and regression model learning in 3D image quality prediction.

For a given set of N arbitrary training samples (F_i, y_i), where F_i represents the quality-predictive features of the ith pair of original/distorted images and y_i is the corresponding subjective quality score, our goal is to find a function which minimizes the deviation from the subjective quality score for all the training data. The function $f(F_i)$ with L hidden nodes can be mathematically modeled and expressed by

$$f(F_i) = \sum_{j=1}^{L} \beta_j \cdot g_j(F_i) = g(F_i) \cdot \beta, \quad i = 1, \ldots, N, \tag{14}$$

where $g(F_i) = [g_1(F_i), \ldots, g_L(F_i)]$ is the output vector of the hidden neuron and $\beta = [\beta_1, \ldots, \beta_L]^T$ denotes the output weighting vector between the output node and the hidden layer of L nodes. The activation function $g(F_i)$ can approximate N training samples by minimizing the training error and can be formulated as

$$g_j(F_i) = g\left(w_j \cdot F_i + b_j\right), \tag{15}$$

where w_j is the weighing vector which connects the input layer and the jth hidden node and b_j denotes the corresponding threshold of the hidden node. In equation (14), β is the only parameter to be determined, which leads to fast learning for ELM [49]. For N training samples (F_i, y_i), the mathematical model for ELM (equation (14)) can be described as follows:

$$\mathbf{Y_H}\beta = \mathbf{Y}, \tag{16}$$

where $\mathbf{Y}$ represents the target vector and $\mathbf{Y_H}$ is called the hidden layer output matrix, which can be defined as

$$\begin{aligned} \mathbf{Y_H} &= \begin{Bmatrix} g(w_1 \cdot F_1 + b_1) & \ldots & g(w_L \cdot F_1 + b_L) \\ \vdots & \vdots & \vdots \\ g(w_1 \cdot F_N + b_1) & \ldots & g(w_L \cdot F_N + b_L) \end{Bmatrix}, \\ \beta &= \begin{Bmatrix} \beta_1 \\ \vdots \\ \beta_L \end{Bmatrix}, \\ \mathbf{Y} &= \begin{Bmatrix} y_1 \\ \vdots \\ y_N \end{Bmatrix}. \end{aligned} \tag{17}$$

The minimal norm least-squares method is used in ELM to minimize the norm of the output weights. Then, the vector of the output weights β can be predicted analytically and expressed by

$$\beta = \mathbf{Y}\mathbf{Y}_\mathbf{H}^\dagger, \tag{18}$$

where $\mathbf{Y}_\mathbf{H}^\dagger$ denotes the Moore–Penrose (MP) generalized pseudoinverse of the hidden layer output matrix $\mathbf{Y_H}$. In practice, the orthogonal projection method [49] can be efficiently employed to calculate the Moore–Penrose inverse:

$$\mathbf{Y}_\mathbf{H}^\dagger = \begin{cases} \left(Y_H^T Y_H\right)^{-1} Y_H^T, & \text{if } Y_H^T Y_H \text{ is nonsingular,} \\ Y_H^T \left(Y_H^T Y_H\right)^{-1}, & \text{if } Y_H Y_H^T \text{ is nonsingular.} \end{cases} \tag{19}$$

Based on the ridge regression theory, a positive value $1/\lambda$ is added to the diagonal of $\mathbf{Y_H^T Y_H}$ or $\mathbf{Y_H Y_H^T}$, which makes the solution more stable. Therefore, with this positive value $1/\lambda$, we can obtain

$$\beta = \begin{cases} \mathbf{Y_H^T}\left(\dfrac{I}{\lambda} + Y_H Y_H^T\right)^{-1} Y, & \text{if } N \le L, \\ \left(\dfrac{I}{(\lambda + Y_H Y_H^T)}\right)^{-1} Y_H^T Y, & \text{if } N > L, \end{cases} \tag{20}$$

where N represents the number of training samples and L denotes the number of hidden nodes. In this paper, the number of nodes L is selected to be equal to the number of training samples N. As a result, the output weight vector β is determined as $\mathbf{Y_H^T}((I/\lambda) + Y_H Y_H^T)^{-1} Y$ in the experiments. More details on the ELM can be found in [49].

4. Experimental Results and Analysis

In the experiments, we first describe the databases and criteria used for quality assessment. Then, we give the performance comparison with other related algorithms in terms of predicting the quality of distorted stereoscopic

images. Moreover, we show the evaluation results on individual distortion type. In addition, we investigate the effect of each component in the proposed metric. Finally, we perform the cross-database evaluation and analyze the time complexity in our experiments.

4.1. Experimental Databases and Protocols. In order to verify and compare the performance of our proposed quality assessment metric, three public and subject-rated benchmark 3D image databases were used as standards: LIVE 3D IQA database Phase I [52], LIVE 3D IQA database Phase II [20], and MCL-3D database [53].

(1) LIVE 3D IQA database Phase I [52]: phase I contains 20 reference stereopairs and 365 symmetrically distorted stereopairs corresponding to five distortion types: JPEG compression, JPEG2000 (JP2K) compression, additive white noise (WN), Gaussian blur (GB), and a simulated fast-fading (FF) model. Each distorted stimulus has been evaluated by human observers and assigned a difference mean opinion score (DMOS) value. The lower DMOS values represent higher visual quality.

(2) LIVE 3D IQA database Phase II [20]: phase II has 120 symmetrically distorted stimuli and 240 asymmetrically distorted stimuli generated from 8 pristine stereopairs. Each of the five distortion types (JPEG, JP2K, WN, GB, and FF) is symmetrically and asymmetrically applied to the pristine stereopairs at various degradation levels. The corresponding DMOS values are also given for the distorted stereopairs.

(3) The MCL-3D database [53]: this database consists of 684 stereoscopic image pairs. Nine image-plus-depth sources are selected, and then a depth-image-based rendering technique is used to render 3D images. Four levels of distortions are applied to either the depth map or texture stereoscopic image prior to 3D image rendering. The distortion types are JPEG, JP2K, WN, Gaussian blur (GBLUR), downsampling blur (SBLUR), and transmission error (TERROR). Each distorted stimulus has been scored by human observers, and a pairwise comparison is used to obtain reliable mean opinion score (MOS) values.

To benchmark the performance of quality assessment metrics, three general performance indicators were employed to provide quantitative performance evaluations: (1) Pearson's linear correlation coefficient (PLCC), which measures the linear dependence between the predicted quality scores and the ground truth targets, (2) Spearman's rank-order correlation coefficient (SRCC), which serves as a measure of prediction monotonicity, and (3) Kendall's rank-order correlation coefficient (KRCC), which is a nonparametric rank-order-based correlation metric. Higher values of PLCC, SRCC, and KRCC represent good consistency with human perceptual quality ratings. For the nonlinear regression, a five-parameter logistic function [54] was applied to fit the predicted quality scores and provided quality scores.

In the experiments, we randomly split each database into two nonoverlapping subsets: a training subset and a test subset. A training process was required to calibrate the quality prediction model. In each train-test procedure, 80% of the database content was selected for training and the remaining 20% for test. After learning the statistical regression model using the training set, the quality prediction performance was evaluated using the test set. In specific, to avoid potential performance bias of the proposed scheme, the train-test iteration was repeated 1000 times, and the median values of PLCC, SRCC, and KRCC were chosen as the final validation results for performance evaluation. In the implementation, a unipolar sigmoidal function $(1/(1+e^{-\lambda u}))$ with $\lambda = 0.1$ was used as the ELM nonlinear activation function.

4.2. Overall Performance Comparison. To comprehensively investigate the effectiveness and robustness of the proposed scheme, we have conducted several different experiments to compare our scheme with the typical and representative methods. These mainly include two 2D-IQA methods (PSNR and multiscale structural similarity (MS-SSIM) [55]), two FR 3D-IQA methods (Benoit et al.'s method [13] and Chen et al.'s method [8]), and three NR 3D-IQA methods (Zhou and Yu's method [56], Fan et al.'s method [41], and Shen et al.'s method [22]). For the previous two 2D-IQA approaches, the predicted quality score of a stereoscopic image was obtained by averaging the left and right image qualities. For Benoit et al.'s approach [13], the disparity distortion was the global correlation between the original and distorted disparity maps. For Chen et al.'s approach [8], we adopted the cyclopean metric in terms of multiscale SSIM described in their paper.

Figure 4 provides the scatter plots of predicted quality scores against subjective DMOS values for the proposed scheme and other compared methods on the LIVE 3D IQA database Phase I. In these figures, the horizontal axis represents the predicted quality scores and the vertical axis denotes the subjective DMOS values of the perceived distortions. Considering performance comparison, a straight-lined distribution of scatter points is better than other arbitrary shapes. For the PSNR and MS-SSIM [55] approaches, the performance is worse than most of other methods in general. The reason can be attributed to that these methods treat the left- and right-view images independently and binocular visual characteristics are not taken into account. For Benoit et al.'s approach [13], the quality evaluation accuracy is even lower than the 2D-IQA approaches under some distortions. One possible explanation is that the 2D image quality metric for disparity maps does not coincide with the human perception of disparity. Overall, the proposed scheme has better consistent alignment with human subjective judgements for stereoscopic 3D images on the database.

In order to further evaluate the performance comparison of quality assessment accuracy on the three databases, we

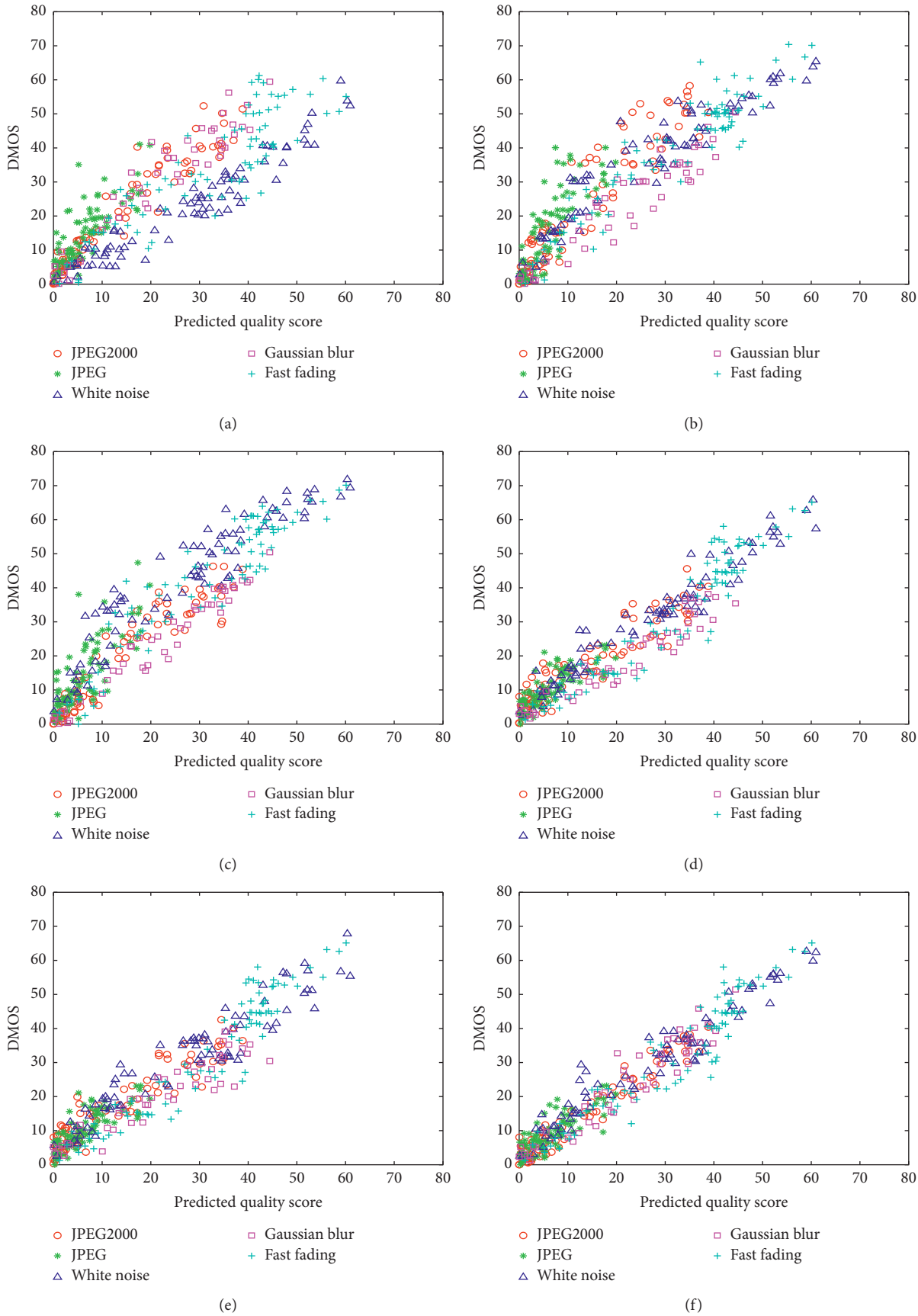


Figure 4: Continued.

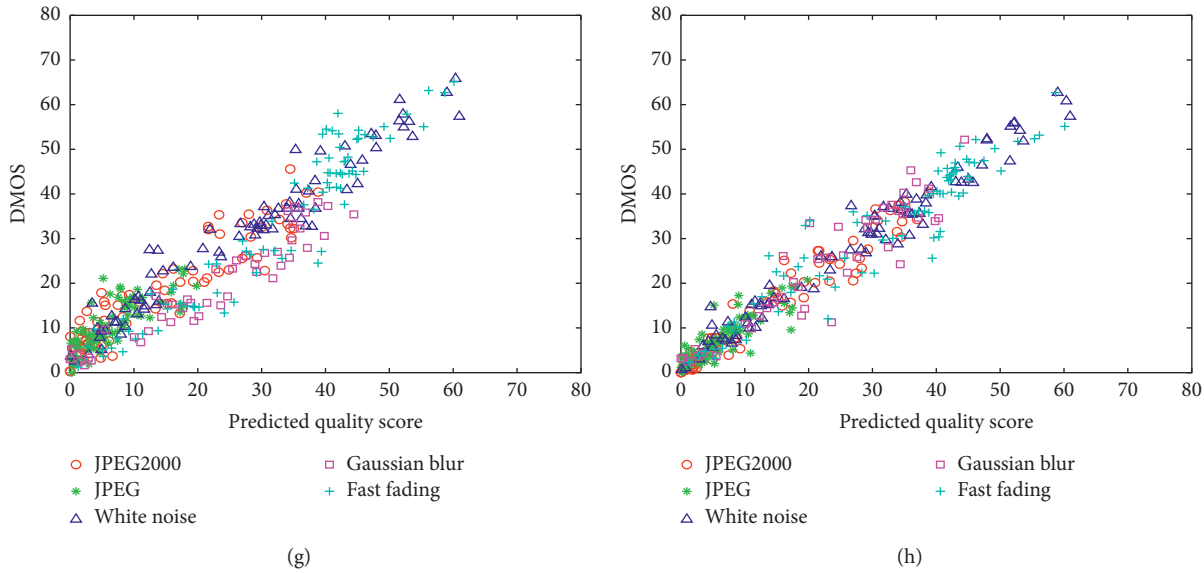


Figure 4: Scatter plots of the predicted quality scores against the subjective DMOS values for the eight methods. (a) PSNR. (b) MS-SSIM [55]. (c) Benoit et al.'s method [13]. (d) Chen et al.'s method [8]. (e) Zhou and Yu's method [56]. (f) Fan et al.'s method [41]. (g) Shen et al.'s method [22]. (h) Proposed scheme.

have given the values of PLCC, SRCC, and KRCC between the provided and predicted quality scores for the proposed scheme and the compared methods. Table 1 presents the performance comparison results in terms of PLCC, SRCC, and KRCC on the three databases. In each case, the results of the best-performance metric are marked in bold. According to the experimental results in this table, Shen et al.'s method [22] performs best on asymmetrically distorted stereoscopic images in the LIVE 3D IQA database Phase II, and our proposed scheme achieves higher consistency with human opinion scores on the other two databases. Moreover, the PLCC and SRCC values for our scheme are above 0.912 and 0.907, respectively, on all databases, which demonstrate that the proposed scheme exhibits a good stability to quantify and predict the perceptual distortions of 3D images. On the whole, the proposed scheme has competitive performance and shows statistically superiority over other typical and representative methods for 3D image quality prediction.

4.3. Distortion-Specific Performance Evaluation. In this section, we have investigated the distortion-specific performance of the proposed scheme and other compared methods for each individual distortion type on the hybrid distortion databases. The PLCC, SRCC, and KRCC comparison results are summarized in Tables 2–4, respectively. For reasons of space and for brevity, M[n] is used to represent the corresponding compared method proposed in paper [n]. The top two quality assessment metrics for each index (PLCC, SRCC, or KRCC) have been highlighted in bold. From these tables, we can find that our proposed metric achieves the highest hit-count for each index and is statistically superior to the compared methods. Some metrics may have high assessment accuracies for specific distortion types: Chen et al.'s method [8] shows strong competitiveness on Gaussian blur, and Shen et al.'s method [22] has outstanding performance on JPEG compression . But our method is comparable to the best-performing metrics for these kinds of distortions. The proposed scheme generally outperforms the vast majority of compared methods by a certain margin for distortion-specific performance evaluation. From these experimental results, it is worth noting that the quality prediction of our scheme is basically independent of different sorts of distortions.

4.4. Contribution of Each Component in the Proposed Scheme. In this section, to understand the respective contributions of each component to the overall quality score in the proposed metric, we have devised three different schemes for comparison, denoted by scheme A, scheme B, and scheme C, respectively. For scheme A, the binocular features of GM response and the binocular energy were used to measure visual quality. For scheme B, the binocular features of LOG response and the binocular energy were adopted for quality prediction. For scheme C, the binocular energy was not included, and the binocular features of GM and LOG responses were considered for quality evaluation. The PLCC, SRCC, and KRCC results are reported in Table 5. As can be observed from this table, the binocular features of LOG response have the most important impact on quality prediction under distortions. It can be inferred that the binocular features of GM and LOG responses and the binocular energy are complementary, and only adopting one aspect of these features cannot obtain the best performance. In addition, according to the results in the table, scheme B has higher assessment accuracies than scheme A, which implies that the LOG features contain more useful visual information and contribute more to 3D quality prediction than the GM features. The experimental results have also

Table 1: Performance of the proposed scheme and other methods in terms of PLCC, SRCC, and KRCC, using the three databases.

IQA model	LIVE-I			LIVE-II			MCL-3D		
	PL	SR	KR	PL	SR	KR	PL	SR	KR
PSNR	0.852	0.846	0.752	0.809	0.790	0.685	0.837	0.816	0.741
MS-SSIM [55]	0.876	0.872	0.815	0.831	0.839	0.727	0.886	0.891	0.759
Benoit [13]	0.883	0.855	0.776	0.835	0.794	0.715	0.843	0.819	0.746
Chen [8]	0.917	0.916	0.737	0.900	0.889	0.730	0.881	0.884	0.717
Chen [17]	0.939	0.923	0.830	0.865	0.859	0.806	0.903	0.892	0.839
Zhou [56]	0.910	0.901	0.711	0.771	0.770	0.550	0.833	0.831	0.612
Fan [41]	0.928	0.887	0.724	0.861	0.823	0.627	0.865	0.830	0.751
Chen [34]	0.937	0.920	0.826	0.870	0.852	0.729	0.907	0.893	0.765
Shen [22]	0.932	0.925	0.839	**0.928**	**0.920**	0.856	0.930	0.922	0.835
Proposed	**0.941**	**0.927**	**0.861**	0.912	0.907	**0.858**	**0.937**	**0.926**	**0.852**

PL: PLCC, SR: SRCC, KR: KRCC. The results of the best-performance metric are marked in bold.

Table 2: PLCC comparison for each type of distortion.

Database	Distortion	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	JPEG	0.521	0.735	0.633	0.837	0.715	0.825	**0.840**	**0.865**
	JP2K	0.826	**0.917**	**0.930**	0.916	0.893	0.909	0.912	0.873
	WN	0.895	0.923	0.919	0.943	0.916	**0.947**	0.935	**0.952**
	GBLUR	0.913	0.936	0.924	**0.951**	0.912	0.950	0.942	**0.958**
	FF	0.740	0.861	0.893	0.915	0.878	0.882	**0.918**	**0.916**
LIVE-II	JPEG	0.726	**0.865**	0.694	0.820	0.770	0.692	**0.871**	0.863
	JP2K	0.783	0.868	0.826	**0.872**	0.611	0.735	**0.904**	0.855
	WN	0.905	**0.940**	0.743	0.902	0.796	0.904	0.920	**0.927**
	GBLUR	0.817	0.793	0.896	**0.965**	0.872	0.913	0.943	**0.946**
	FF	0.802	0.816	0.851	0.909	0.904	0.851	**0.929**	**0.932**
MCL-3D	JPEG	0.651	**0.870**	0.674	0.861	0.805	0.856	0.859	**0.864**
	JP2K	0.782	0.879	0.906	**0.913**	0.872	**0.910**	0.909	0.885
	WN	0.886	**0.955**	0.792	0.892	0.856	0.866	0.935	**0.939**
	GBLUR	0.867	0.914	0.839	0.867	0.757	0.915	**0.951**	**0.954**
	SBLUR	0.829	0.892	0.815	0.907	0.819	**0.924**	**0.926**	0.920
	TERROR	0.813	0.867	0.790	**0.905**	0.743	0.769	0.893	**0.927**

The top two quality assessment metrics have been highlighted in bold.

demonstrated that the quality assessment performance can be promoted by appropriate combination of binocular features and binocular energy.

4.5. Cross-Database Performance Evaluation. In the above experiments, the training and test subset have the same distortions selected from the databases. Since the proposed scheme is based on a learning framework, it is necessary to ascertain whether the performance is bound to a special training database on which it is trained. To verify the generalization ability and stability of our scheme, we have carried out cross-database experiments for performance evaluation. In the experiment, we examined whether satisfactory results could be obtained by applying the regression model trained on one database to the testing set from another database. For brevity, the SRCC results of cross-database performance evaluation are given in Table 6. It can be observed that the proposed metric has comparatively weak performance in comparison with the evaluation results in Table 3. The reason can be mainly attributed to that the training and test subsets have different types of distortions. For instance, the LIVE Phase I database only has symmetrically distorted images, while the LIVE Phase II database contains both symmetric and asymmetric distortions. However, the values of corresponding indicators are still relatively high, which show that our framework can maintain a satisfying predictive capacity under different circumstances. Based on the above experimental results, it can be concluded that a larger training database with more comprehensive distortion types could probably promote the prediction accuracy of our scheme.

In this section, we have also compared the cross-database evaluation of the proposed scheme with other related methods. The SRCC results are provided in Table 7, where the top two metrics have been marked in bold. According to the experimental results in the table, no matter which training database is used, the cross-database evaluation of our scheme is usually stable, and it offers statistically better quality prediction in line with human perception than the compared methods on different databases. These facts demonstrate the generalization ability and effectiveness of the proposed scheme for stereoscopic image quality assessment.

Table 3: SRCC comparison for each type of distortion.

Database	Distortion	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	JPEG	0.507	0.725	0.652	0.816	0.569	0.614	**0.845**	**0.842**
	JP2K	0.818	**0.906**	0.897	0.877	0.812	0.824	**0.909**	0.853
	WN	0.902	0.925	0.912	0.895	**0.940**	0.915	0.937	**0.942**
	GBLUR	0.895	0.933	0.874	**0.953**	0.860	0.916	0.938	**0.963**
	FF	0.754	0.856	0.860	**0.912**	0.784	0.867	0.910	**0.924**
LIVE-II	JPEG	0.696	**0.852**	0.704	0.822	0.769	0.593	0.843	**0.847**
	JP2K	0.785	0.847	0.819	**0.863**	0.593	0.717	**0.886**	0.829
	WN	0.873	0.934	0.738	0.889	0.846	0.891	**0.936**	**0.939**
	GBLUR	0.809	0.803	0.904	**0.956**	0.862	0.903	**0.951**	0.932
	FF	0.806	0.821	0.836	0.895	**0.935**	0.891	0.901	**0.909**
MCL-3D	JPEG	0.645	0.861	0.719	0.857	0.781	0.848	**0.863**	**0.868**
	JP2K	0.809	0.873	0.843	0.902	0.875	**0.927**	**0.908**	0.897
	WN	0.891	**0.942**	0.758	0.899	0.823	0.861	0.912	**0.932**
	GBLUR	0.864	0.906	0.827	0.893	0.705	0.910	**0.926**	**0.945**
	SBLUR	0.832	0.880	0.796	0.920	0.818	**0.951**	**0.933**	0.904
	TERROR	0.815	0.852	0.807	**0.903**	0.609	0.676	0.895	**0.907**

The top two quality assessment metrics have been highlighted in bold.

Table 4: KRCC comparison for each type of distortion.

Database	Distortion	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	JPEG	0.434	0.692	0.558	0.736	0.535	0.601	**0.757**	**0.826**
	JP2K	0.685	**0.826**	0.804	0.727	0.793	0.706	**0.859**	0.803
	WN	0.823	0.829	0.811	0.736	**0.912**	0.835	0.820	**0.915**
	GBLUR	0.806	0.850	0.786	**0.863**	0.839	0.827	0.858	**0.931**
	FF	0.638	0.729	0.752	0.856	0.706	0.772	**0.863**	**0.882**
LIVE-II	JPEG	0.567	0.735	0.591	0.727	0.522	0.571	**0.749**	**0.736**
	JP2K	0.632	**0.783**	0.754	0.726	0.537	**0.765**	0.736	0.725
	WN	0.729	0.821	0.695	0.715	0.696	0.619	**0.833**	**0.839**
	GBLUR	0.656	0.806	0.827	**0.862**	0.729	0.820	0.881	**0.894**
	FF	0.694	0.715	0.709	0.739	**0.853**	0.783	**0.872**	0.851
MCL-3D	JPEG	0.656	0.719	0.682	0.709	0.607	0.552	**0.753**	**0.772**
	JP2K	0.727	0.753	0.817	0.763	0.715	**0.863**	0.827	**0.846**
	WN	0.825	**0.851**	0.786	0.685	0.616	0.639	0.849	**0.895**
	GBLUR	0.796	0.764	0.624	0.690	0.596	0.797	**0.886**	**0.904**
	SBLUR	0.809	0.766	0.732	**0.817**	0.775	0.812	**0.825**	0.813
	TERROR	0.751	0.705	0.724	**0.761**	0.579	0.537	0.758	**0.819**

The top two quality assessment metrics have been highlighted in bold.

Table 5: Performance of each component in the proposed scheme.

	LIVE-I			LIVE-II			MCL-3D		
	PL	SR	KR	PL	SR	KR	PL	SR	KR
Scheme A	0.832	0.754	0.681	0.724	0.658	0.593	0.761	0.696	0.605
Scheme B	0.879	0.826	0.735	0.803	0.717	0.642	0.829	0.740	0.652
Scheme C	0.851	0.797	0.714	0.750	0.664	0.607	0.785	0.728	0.637
Proposed	0.941	0.927	0.861	0.912	0.907	0.858	0.937	0.926	0.852

PL: PLCC, SR: SRCC, KR: KRCC.

4.6. Time Complexity Analysis. Time complexity is a significant indicator in evaluating the performance of the proposed scheme, to facilitate its use in real-time applications such as monitoring and adjustment. We have compared the computational complexity of our proposed scheme with other related methods. The experiment was performed in MATLAB R2014a on a Windows 10 PC with a 2.5 GHz Intel Core i7 processor and 8 GB RAM. The results of time consumption are given in Table 8, which presents the running time comparison on the LIVE Phase I database with 365 stereopairs. As can be seen from this table, the total processing time of the proposed scheme is 127 seconds, which indicates that it takes less than 0.35 seconds to predict a distorted stereopair. Although it is not the most efficient

Table 6: Cross-database performance evaluation in terms of SRCC.

Testing database	Training database	JPEG	JP2K	WN	GBLUR	FF	All
LIVE-I	LIVE-II	0.834	0.833	0.896	0.925	0.908	0.904
	MCL-3D	0.821	0.839	0.908	0.885	—	0.890
LIVE-II	LIVE-I	0.815	0.801	0.892	0.907	0.903	0.840
	MCL-3D	0.833	0.754	0.898	0.902	—	0.879
MCL-3D	LIVE-I	0.819	0.868	0.917	0.902	—	0.875
	LIVE-II	0.861	0.870	0.906	0.922	—	0.905

Table 7: SRCC comparison with other related methods for cross-database performance evaluation.

Testing	Training	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
LIVE-I	LIVE-II	0.792	0.839	0.817	0.896	0.871	0.866	**0.901**	**0.904**
	MCL-3D	0.808	0.865	0.853	**0.913**	**0.897**	0.885	0.895	0.890
LIVE-II	LIVE-I	0.729	0.835	0.762	**0.861**	0.765	0.816	**0.857**	0.840
	MCL-3D	0.746	0.811	0.783	0.877	0.752	0.820	**0.883**	**0.879**
MCL-3D	LIVE-I	0.747	**0.861**	0.710	0.853	0.812	0.825	0.859	**0.875**
	LIVE-II	0.803	0.823	0.815	0.817	0.819	0.828	**0.868**	**0.905**

The top two metrics have been marked in bold.

Table 8: Performance comparison of time complexity on the LIVE Phase I database.

Metrics	PSNR	M [55]	M [13]	M [8]	M [56]	M [41]	M [22]	Proposed
Run time (s)	7105	9520	6814	5143	37	453	396	127

method, it has the best comprehensive performance in achieving the balance between accuracy and timeliness. The simulation results demonstrate that our proposed scheme has relatively lower computing complexity than the compared methods.

5. Conclusions

In this paper, we have presented a novel no-reference quality prediction method for stereoscopic images based on binocular statistical features and machine learning. The framework of the proposed scheme includes a feature extraction stage and a feature mapping stage. The gradient magnitude and Laplacian of Gaussian responses from a stereopair and its fusion and difference maps are utilized as quality-predictive features. With the extreme learning machine, a statistical regression model is established to map these binocular features of a stereopair to its corresponding perceptual quality score. The visual quality predictions by the proposed metric are highly correlated with subjective quality judgements for distorted image pairs of various distortion types. More importantly, our method achieves excellent performance and has a promising generalization ability. The proposed scheme can be applied to video broadcasting and 3D multimedia industry for its practicality.

For future work, how to explore deeper structure representation for a human visual system and how to design more efficient machine learning methods for visual quality prediction should be researched. In addition, more effective quality features can be considered to simulate the human perceptual vision. Other 3D quality factors such as depth perception and visual comfort still deserve further study.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported in part by the Key Research and Development Program of China under Grant no. 2018YFC0831000, the National Natural Science Foundation of China under Grant no. 62001267, the Natural Science Foundation of Shandong Province under Grant no. ZR2020QF013, the Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project) under Grant no. 2019JZZY010119, and the Fundamental Research Funds of Shandong University under no. 2020HW017.

References

[1] L. Chen and J. Zhao, "Quality assessment of stereoscopic 3D images based on local and global visual characteristics," in *Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 61–66, Hong Kong, China, July 2017.

[2] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: visual comfort," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 335–346, 2011.

[3] S. Khan and S. S. Channappayya, "Estimating depth-salient edges and its application to stereoscopic image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5892–5903, 2018.

[4] B. Appina, S. V. R. Dendi, K. Manasa, S. S. Channappayya, and A. C. Bovik, "Study of subjective quality and objective blind quality prediction of stereoscopic videos," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5027–5040, 2019.

[5] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[6] X. Ben, C. Gong, P. Zhang, X. Jia, Q. Wu, and W. Meng, "Coupled patch alignment for matching cross-view gaits," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3142–3157, 2019.

[7] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[8] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[9] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, 2013.

[10] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: a natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[11] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2019.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[13] A. Benoit, P. L. Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Proceedings of the 15th IEEE International Conference on Image Processing*, pp. 389–392, San Diego, CA, USA, October 2008.

[14] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," in *Proceedings of the Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM2010)*, pp. 61–66, Scottsdale, AZ, USA, November 2010.

[15] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1527–1542, 2014.

[16] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1940–1953, 2013.

[17] L. Chen and J. Zhao, "Perceptual quality assessment of stereoscopic images based on local and global visual characteristics," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 12139–12156, 2019.

[18] X. Ben, P. Zhang, Z. Lai, R. Yan, X. Zhai, and W. Meng, "A general tensor representation framework for cross-view gait recognition," *Pattern Recognition*, vol. 90, pp. 87–98, 2019.

[19] X. Ben, C. Gong, P. Zhang, R. Yan, Q. Wu, and W. Meng, "Coupled bilinear discriminant projection for cross-view gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 734–747, 2020.

[20] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, 2013.

[21] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 591–602, 2014.

[22] L. Shen, R. Fang, Y. Yao, X. Geng, and D. Wu, "No-reference stereoscopic image quality assessment based on image distortion and stereo perceptual information," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 1, pp. 59–72, 2019.

[23] L. Chen and J. Zhao, "No-reference quality assessment for stereoscopic 3D images based on binocular visual perception," in *Proceedings of the IEEE International Symposium on Haptic, Audio and Visual Environments and Games*, pp. 1–5, Dalian, China, September 2018.

[24] J. Yang, C. Ji, B. Jiang, W. Lu, and Q. Meng, "No reference quality assessment of stereo video based on saliency and sparsity," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 341–353, 2018.

[25] Z. Chen, J. Xu, C. Lin, and W. Zhou, "Stereoscopic omnidirectional image quality assessment based on predictive coding theory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 103–117, 2020.

[26] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[27] J. Wu, S. Pan, X. Zhu, and Z. Cai, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.

[28] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," in *Proceedings of the International Joint Conference on Artificial Intelligence*, Yokohama, Japan, August 2020.

[29] W. Zhou, W. Qiu, and M.-W. Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of 3-D images," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 404–415, 2017.

[30] S. Li, X. Han, and Y. Chang, "Adaptive cyclopean image-based stereoscopic image-quality assessment using ensemble learning," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2616–2624, 2019.

[31] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1733–1740, Columbus, OH, USA, June 2014.

[32] H. Oh, S. Ahn, J. Kim, and S. Lee, "Blind deep S3D image quality evaluation via local to global feature aggregation," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4923–4936, 2017.

[33] W. Zhou, Z. Chen, and W. Li, "Dual-stream interactive networks for no-reference stereoscopic image quality

assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 3946–3958, 2019.

[34] L. Chen and J. Zhao, "No-reference perceptual quality assessment of stereoscopic images based on binocular visual characteristics," *Signal Processing: Image Communication*, vol. 76, pp. 1–10, 2019.

[35] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Blind quality estimator for 3D images based on binocular combination and extreme learning machine," *Pattern Recognition*, vol. 71, pp. 207–217, 2017.

[36] D. Stidwill and R. Fletcher, *Normal Binocular Vision: Theory, Investigation and Partical Aspects*, Wiley, New York, NY, USA.

[37] N. Kruger, P. Janssen, S. Kalkan et al., "Deep hierarchies in the primate visual cortex: what can we learn for computer vision?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1847–1871, 2013.

[38] J. Wu, W. Lin, and G. Shi, "Image quality assessment with degradation on spatial structure," *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 437–440, 2014.

[39] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.

[40] K. A. May and L. Zhaoping, "Efficient coding theory predicts a tilt aftereffect from viewing untilted patterns," *Current Biology*, vol. 26, no. 12, pp. 1571–1576, 2016.

[41] Y. Fan, M. C. Larabi, F. A. Cheikh, and C. F. Maloigne, "No-reference quality assessment of stereoscopic images based on binocular combination of local features statistics," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 3538–3542, Athens, Greece, October 2018.

[42] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images," *ACM Transactions on Applied Perception*, vol. 3, no. 2, pp. 95–109, 2006.

[43] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.

[44] K. Lee and S. Lee, "3D perception based quality pooling: stereopsis, binocular rivalry, and binocular suppression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 533–545, 2015.

[45] Q. Peng and B. E. Shi, "The changing disparity energy model," *Vision Research*, vol. 50, no. 2, pp. 181–192, 2010.

[46] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 20, no. 8, pp. 2378–2386, 2011.

[47] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[48] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[49] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[50] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 42, no. 2, pp. 513–529, 2012.

[51] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "NMF-based image quality assessment using extreme learning machine," *IEEE Transactions on Cybernetics*, vol. 47, pp. 232–243, 2017.

[52] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process: Image Communication*, vol. 28, no. 8, pp. 870–883, 2012.

[53] R. Song, H. Ko, and C. C. J. Kuo, "MCL-3D: a database for stereoscopic image quality assessment using 2D-image-plus-depth source," *Journal of Information Science and Engineering*, vol. 31, no. 5, pp. 1593–1611, 2015.

[54] P. G. Gottschalk and J. R. Dunn, "The five-parameter logistic: a characterization and comparison with the four-parameter logistic," *Analytical Biochemistry*, vol. 343, no. 1, pp. 54–65, 2005.

[55] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proceedings of the IEEE The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, pp. 1398–1402, Pacific Grove, CA, USA, November 2003.

[56] W. Zhou and L. Yu, "Binocular responses for no-reference 3D image quality assessment," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1077–1084, 2016.

Hindawi
Complexity
Volume 2021, Article ID 6610965, 8 pages
https://doi.org/10.1155/2021/6610965

Research Article

A Novel Chinese Entity Relationship Extraction Method Based on the Bidirectional Maximum Entropy Markov Model

Chengyao Lv,[1] **Deng Pan**,[2] **Yaxiong Li**,[3] **Jianxin Li**,[4] **and Zong Wang**[5]

[1]*School of Foreign Languages, China University of Geosciences, Wuhan 430074, China*
[2]*School of Foreign Languages, Hubei University of Science and Technology, Xianning 437000, China*
[3]*Information Centre, Hubei University of Science and Technology, Xianning 437000, China*
[4]*School of Information Technology, Deakin University, Geelong VIC3220, Australia*
[5]*School of Public Administration, China University of Geosciences, Wuhan 430074, China*

Correspondence should be addressed to Yaxiong Li; liyaxiong@hbust.edu.cn

Received 9 October 2020; Revised 14 December 2020; Accepted 5 January 2021; Published 19 January 2021

Academic Editor: Jia Wu

To identify relationships among entities in natural language texts, extraction of entity relationships technically provides a fundamental support for knowledge graph, intelligent information retrieval, and semantic analysis, promotes the construction of knowledge bases, and improves efficiency of searching and semantic analysis. Traditional methods of relationship extraction, either those proposed at the earlier times or those based on traditional machine learning and deep learning, have focused on keeping relationships and entities in their own silos: extracting relationships and entities are conducted in steps before obtaining the mappings. To address this problem, a novel Chinese relationship extraction method is proposed in this paper. Firstly, the triple is treated as an entity relation chain and can identify the entity before the relationship and predict its corresponding relationship and the entity after the relationship. Secondly, the Joint Extraction of Entity Mentions and Relations model is based on the Bidirectional Long Short-Term Memory and Maximum Entropy Markov Model (Bi-MEMM). Experimental results indicate that the proposed model can achieve a precision of 79.2% which is much higher than that of traditional models.

1. Introduction

In the age of big data, techniques of extracting valuable information from enormous quantities of texts have drawn the attention of many researchers. The extraction of information includes entity extraction, relationship extraction, and event extraction. As the key step in information extraction, relationship extraction provides technical foundation for subsequent tasks such as knowledge graphs, intelligent information retrieval, and semantic analysis. Therefore, techniques of relationship extraction are beneficial not only for theoretical discussion but also for practical application.

Research on techniques to extract entities and their relationships can date back to the 1960s. Among the more prominent projects is the Linguistic String Project by New York University, which took the route of constructing massive language (English) corpora and achieved very satisfactory results when the team used these corpora to extract information from medical texts. In addition, a systematic research at Yale University extracted events in domains such as "earthquake" and "strike" from news texts and promoted the research and development of entity relationship extraction. By the late 1980s, with the convening of the Message Understanding Conference, research on entity relationship extraction had started to boom. After decades' development, theories and techniques of entity relationship extraction, from early models of manual design and rule extraction [1, 2] to late models based on machine learning [3] and deep learning [4, 5], are approaching maturity. With constant improvements in model accuracy and recall, extraction models are more adaptive than ever before.

However, the most existing extraction techniques either have been keeping relationships and entities in their own silos. Extracting relationships and entities was conducted in separate steps before obtaining the mappings, or tag triples as a whole used the "proximity principle" of reinforced learning to extract relationships. Existing extraction techniques fit into three categories. Firstly, the relationship can be predicted and identified by an entity pair. The premise of this idea is that the relationships are already predefined [6]. The task of relationship extraction then becomes the task of searching the predefined relationship space for the most probable relationship between a given entity pair based the context where the entity pair is located. Secondly texts can be explored by the relationship of entity pairs. This method aims at finding the maximum number of entity pairs matching the criteria of the given relationship. A common issue of the two methods mentioned above is the subtasks, entity identification and relationship identification, are completely independent of each other, resulting in extraneous information such as entities without relationship. This, in turn, increases error rates because the entities are paired up before their relationship is determined; when no relationship is found for an entity pair, this pair becomes extraneous. Such extraneous pairs increase error rates of the subtask and negatively impact the performance of subsequent relationship classification. Finally, some studies tag triples as a whole and use the "proximity principle" of reinforced learning to extract relationships [7]. This method integrates low-level features into more abstract high-level features to search for distributed feature representations and, thus, solves the problems of manual feature selection and the spread of feature extraction error haunting classical methods.

The conventional method has two drawbacks. Firstly, for most of the entity pairs do not hold relationships, numerous negative cases and imbalanced relationship classification occur. Secondly, overlapping triples become a critical issue. The shared entities or multiple relationships between two entities make learning more complicated or even impossible, since adequate training data cannot be obtained. For instance, "Mr. Zhang was born in Hubei, a province in Central China" could be interpreted into <Mr. Zhang, was born in, Hubei>, <Mr. Zhang, was born in, China>, and <Hubei, lies in, China>. The conventional algorithm cannot identify and classify properly without sufficient data.

To address these problems, this paper proposes a new method, entity relation chain. The head entity before relationship should be identified firstly, and then, the corresponding relationship and the tail entity can be predicted. For instance, in the sentence "Mr. Zhang was born in Hubei province," $E1$ "Mr. Zhang" and $E2$ "Hubei province" are usually identified firstly and the R "was born in" is recognized secondly. But, in the entity relation chain, $E1$ "Mr. Zhang" is firstly identified, and every possible R generated from $E1$ is the criterion for $E2$ "Hubei province.' In this entity relation chain, $E1$ can be taken as head entity, R as relation chain, and $E2$ as tail entity.

Experiments on data sets from People's Daily indicated that the proposed method can achieve a high performance. We also evaluated the scalability of the method on English data sets of the English SemEval 2010 Task 8 which reveal that the Bi-MEMM also can obtain a better f-score.

This paper is organized as follows. Starting with the introduction of the research gap and our research purpose, we review and discuss the entity relationship extraction and the particularity of Chinese relation extraction. Then, we develop the Bi-MiEM method for the entity relation extraction. The detailed experimental evaluation is illustrated in Section 4, and Section 5 concludes this work and provides the future direction for further research.

2. Related Work

2.1. Definition of Entity Relationship Extraction. Entity relationship extraction is usually described as entity relationship triples $<E_1, R, E_2>$, in which E_1 and E_2 refer to the entity type and R refers to the relation description type text. After the preprocessing process of named entity recognition relation trigger word recognition, the determined triples $<E_1, R, E_2>$ are stored for further analysis or query.

According to the definition, we can divide the entity relationship extraction tasks into three key parts, name entity recognition, relation trigger word identification, and relation extraction. Name entity recognition refers to the identification of text having a specific meaning of the entity, mainly including the names of people and places, institutions, and proper nouns. Relation trigger word identification is to classify the words that trigger entity relationship, identify whether they are trigger words, and determine whether the extracted relations are positive. Relation extraction is the extraction of semantic relationships between entities from identified entities, such as location employee products.

2.2. Features of Entity Relationship Extraction. Compared with NLP tasks such as sentiment analysis and news classification, the extraction of relationship is unique in three aspects.

Firstly, Entity Relationship Extraction covers diverse domains. Researchers usually focus on one domain or a limited number of domains. With limited relationship categories, traditional techniques are mostly based upon rules [2, 8–10], dictionaries [1, 11], and ontologies [3, 12]. Machine learning-based techniques include supervised [6, 13], semisupervised [14, 15], and unsupervised [16, 17] models. Lately, deep learning-based techniques include supervised [18, 19] and distant supervised [20] models. All these models are relatively easy to build, but with poor portability and extensibility.

Secondly, Entity Relationship Extraction involves heterogenous data. Data can come from different sources, and they can be structured, semistructured, or nonstructured. Deep learning [21] is usually applied in structured data; nonsupervised aggregation methods [4] are usually applied in nonstructured textual data due to unpredictable relationship categories; semisupervised [17] or distant

supervised [22] methods are usually applied in semi-structured data such as Wikipedia.

Lastly, Entity Relationship Extraction needs to handle various relationships, which easily leads to data noise. Relationships between entities are various, but early research often ignored such multiple relationships and failed to handle latent relationships. The adoption of graph structures [18] in relationship extraction in recent years ushered in a new technique for tackling overlaps of entities and relationships. To tackle data noise [23], it has been discovered that using a small number of adversarial examples can avoid model overfitting and proposed to use adversarial training to improve model performance.

2.3. Particularity of Chinese Relation Extraction. Relationship extraction of Chinese texts falls behind the extraction of English because of its complexity and difficulty. The following two characteristics of Chinese make it more challenging for Chinese than English in terms of relationship extraction.

Chinese trigger words are hard to extract and are in abundance. This makes the recall rate of relationship extraction low. In the ACE corpus, Chinese trigger words are 30% more than those in English [24].

For the Chinese language, words are often polysemous, sentence structures are complex and flexible, and omissions appear frequently. The fact that the same word can express completely different meanings in different contexts or the same meaning can be represented with many different expressions makes the identification of relationship types particularly difficult.

In view of these problems, this paper proposes the following possible solutions. Firstly, the Joint Extraction of Entity Mentions and Relations model similar to Seq2Seq is proposed and the Bidirectional Maximum Entropy Markov is integrated into the model. Secondly, different from the existing relationship extraction techniques, relationship triples are treated as an entity relationship chain, entity E_1 is identified first, and then, the corresponding relationship R and entity E_2 based on E_1 are predicted. Thirdly, the validity of the proposed model is verified in Chinese data sets and the scalability is evaluated in English data sets.

3. Extraction Method Based on the Bi-MEMM Model

The previous solutions cannot efficiently deal with the entity relationship extraction entity overlap, relationship crossover, and so on. In this paper, a Bi-MEMM model similar to seq2seq simulated probability graph is proposed to solve such problems. The seq2seq decoder is modeled in the following way:

$$P(y_1, y_2, \ldots, y_n|x) = P(y_1|x)P(y_2|x, y_1)\ldots P \cdot (y_n|x, y_1, y_2, \ldots, y_{n-1}). \quad (1)$$

In formula (1), the first word is predicted by x and the second word is predicted if the first word is known and repeated until the end mark appears. Similarly, the extraction of triples can be modeled in the following way:

$$P(E_1, R, E_2) = P(E_1)P(E_2|E_1)\mathrm{P}(R|E_1t, nE_2). \quad (2)$$

In formula (2), "E_1" can be predicted first, and "E_2" corresponding to "E_1" can be predicted by passing in "E_1". Then, E_1 and E_2 can be introduced to predict the relationship R between E_1 and "E_2." In actual processing, we can also combine the predictions of E_2 and R into one step, so the total step only needs two steps; the first step is to predict E_1, and then, E_1 is introduced to predict E_2 and R corresponding to "E_1."

3.1. Bi-MEMM Model. Figure 1 demonstrates the overall structure of our Bi-MEMM model. It can be detailed as follows.

When it comes to techniques for extracting relationships and entities, character-word embedding is necessary only for Chinese, as word embedding is sufficient for English. By means of word segmentation with Chinese texts, we obtain character embedding and word embedding. Then, we perform matrix transformation of word embedding and concatenate the transformed word embedding with character embedding of the word's constituent characters. The result of such concatenation is character-word embedding. For instance, "中国" has two character-word embedding: one is the concatenation of the matrix-transformed word embedding with character "中", and the other is the concatenation of the matrix-transformed word embedding with character "国".

Firstly, character-word-position embedding is transformed into coding matrix M through the Bi-LSTM Layer and Tanh Layer/Attention Layer.

Secondly, matrix M is copied into the Bi-MEMM Layer and Dense Layer. Sigmoid can be used as activation function for the Dense Layer. Then, a two-dimension vector generated by each character can be used to predict the head and tail position of E1.

Thirdly, a labelled E_1 is randomly picked (randomly pick E_1 when training, and traverse all E_1's when predicting), the subsequence corresponding to E_1 is fed in M into the first Self Attention Layer, together with the Position Embedding at corresponding position, and it transformed into a vector with the same length as the input sequence.

Lastly, matrix M is sent into the Bi-MEMM Layer and Dense Layer again. For each R corresponding to E_1, the head and tail positions of E_2 can be also predicted by the Dense Layer with the activation function of sigmoid.

From the model structure of Figure 1, we can figure out it is similar to the copy mechanism, joint extraction model. In entity '$E1$' identification, <E_1, R, E_2>, Bi-MEMM plays the same role as CRF. In $E2$ recognition, Bi-MEMM predicts $E2$ by every possible R with $E1$. If there is $E2$, the corresponding triples are regarded as an option or the triples will be discarded.

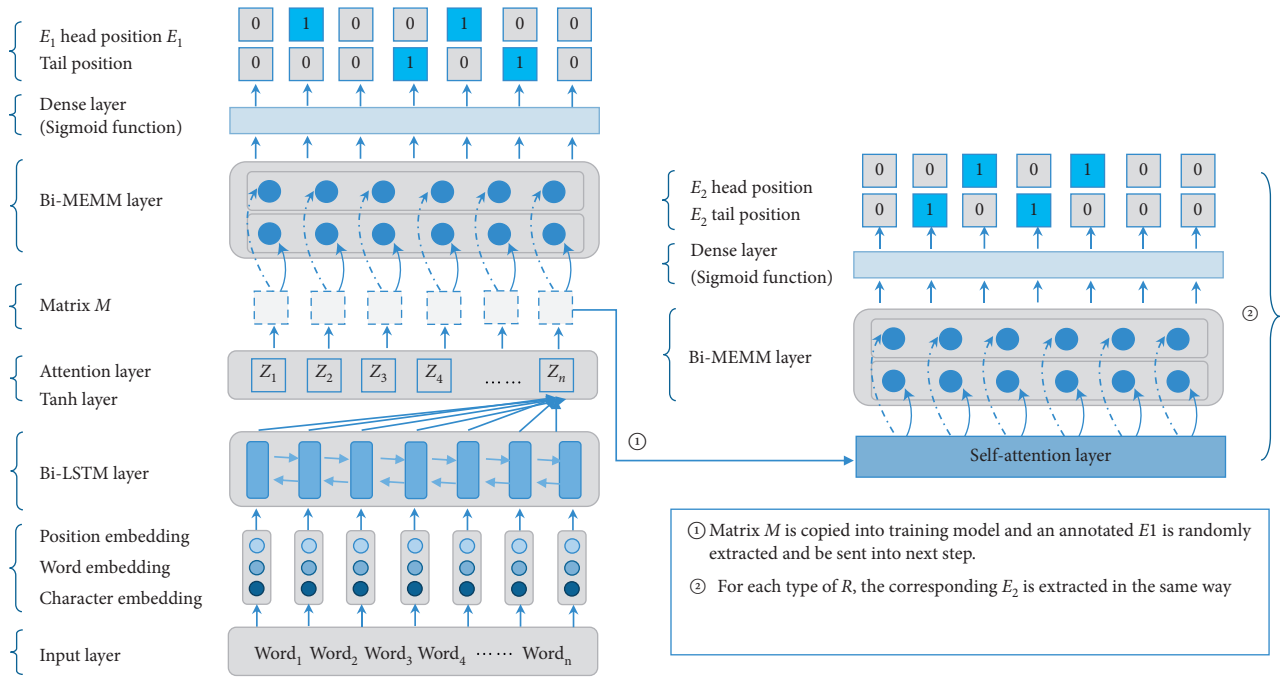


Figure 1: The structure of Bi-MEMM.

3.2. Bi-MEMM Construction and the Loss Function. In formula (1), we assume that the dependency occurs only in adjacent locations, and the following formula is obtained:

$$P(y_1, y_2, \ldots, y_n|x) = P(y_1|x)P(y_2|x, y_1)P(y_3|x, y_2)\ldots P \cdot (y_n|x, \ldots, y_{n-1}). \tag{3}$$

In formula (3), $X = (x_1, x_2, \ldots, x_n)$ is the input and $Y = (y_1, y_2, \ldots, y_n)$ is the tag sequence with the same size of X. According to the design of Linear CRF (Linear Chain Conditional Random Field), the following formula is obtained from formula (3):

$$P(y_1|x) = \frac{e^{f(y_1;x)}}{\sum_{y_1} e^{f(y_k;x)}},$$
$$P(y_k|x, y_{k-1}) = \frac{e^{g(y_{k-1},y_k)+f(y_k;x)}}{\sum_{y_k} e^{g(y_{k-1},y_k)+f(y_k;x)}}, \tag{4}$$

where $g(y_{k-1}, y_k)$ is called the transition matrix. At this point, this is the MEMM. From equation (4), we can see that the solution of the MEMM is to decompose the overall probability distribution into the product of a stepwise distribution, so to calculate the loss, you only need to sum the cross entropy of each step.

Substituting equation (4) into equation (3), we can get the loss of MEMM as follows:

$$P\leftarrow y|x\leftarrow = \frac{e^{f(y_1;x)+g(y_1,y_2)+\ldots+g(y_{n-1},y_n)+f(y_n;x)}}{\left(\sum_{y_1} e^{f(y_1;x)}\right)\left(\sum_{y_2} e^{g(y_1,y_2)+f(y_2;x)}\right)\ldots\left(\sum_{y_n} e^{g(y_{n-1},y_n)+f(y_n;x)}\right)}. \tag{5}$$

So far, we can see that MEMM, like seq2seq, has one significant defect: exposure bias [25]. When the model is trained, the prediction of the current step assumes that the labels of the previous step are correct and acquired. However, in the prediction stage, the actual labels of the previous step are unknown. If the current step is not strengthened during training, the reliability of the entire data chain will be greatly reduced.

The way to calculate the probability of equation (5) is from left to right. Experiments show that adding a right-to-

left MEMM during modelling with reference to the LSTM and Bi-LSTM modes can improve its effect. Then, we can get the following loss function.

Finally, the average cross entropy of formulae (5) and (6) are taken as the final loss. This can make up for the shortcomings of its asymmetric behaviour without increasing the parameters, and it can also strengthen the current training.

Table 1: Hyperparameters settings.

Hyperparameters	Chinese	English
Learning rate	0.0005	0.0005
Hidden neurons size	250	200
Dropout	0.6	0.5
Character embeddings/word embeddings	300	300
Batch size	16	32

$$P\overleftarrow{y|x} = \frac{e^{f(y_1;x)+g(y_1,y_2)+\ldots+g(y_{n-1},y_n)+f(y_n;x)}}{\left(\sum_{y_n} e^{f(y_n;x)}\right)\left(\sum_{y_{n-1}} e^{g(y_n,y_{n-1})+f(y_{n-1};x)}\right)\ldots\left(\sum_{y_1} e^{g(y_2,y_1)+f(y_1;x)}\right)}. \quad (6)$$

4. Experimental Design

Experiments are carried out to evaluate the efficiency of proposed method on Chinese data sets and the scalability on English data sets. For the Chinese data set, corpus data from People's Daily in January in the news field are collected, and the English data set adopted SemEval 2010 Task 8.

Several similar methods such as Bi-LSTM + CRF [5], Att-Bi-LSTM + CRF [26], and bert-based [27] were taken as the base line on the Chinese entity relationship extraction test. The proposed joint extraction model is applied to Chinese data sets to verify its validity. BRNN [28], SDP-BLSTM [29], CNN [30], Att-RCNN [31], and Hybrid Bi-LSTM-Siamese [32] are also carried out as the base line for the scalability evaluation.

4.1. Data Sets. SemEval 2010 Task 8 marks the semantic relationship between noun pairs in a sentence rather than entity pairs. There are 10 classes (cause-effect, component-whole, entity-destination, product-producer, entity-origin, member-collection, message-topic, content-container, instrument-agency, and others) in total, among which one type does not distinguish the sequence of relationship arguments.

The corpus of People's Daily mainly includes three kinds of entity relations, personal name, place name, and organization name. In this paper, Spacy [33], PyhanLP [34], and other natural language processing auxiliary tools [35] are used in experiments.

4.2. Hyperparameters. Due to differences in the data set of Chinese and English, for example, factors Embedding of China Character and Word Embedding of English are not consistent with some superparameters. In this paper, the average cross entropy of formula (6) is used as the loss function to train deep learning network with an Adam optimizer. The superparameters are shown in Table 1.

4.3. Evaluation Criteria. Precision, recall, and F-measure are adopted as the basic evaluation criteria, in which precision and recall are contradictory and F-measure is taken to evaluate comprehensively and globally. Their calculation formulae are listed, respectively, as follows:

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}, \quad (7)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}, \quad (8)$$

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (9)$$

4.4. Experimental Results and Analysis. For the Chinese entity relationship extraction dataset, Bi-LSTM-CRF, Att-Bi-LSTM-CRF, and bert-based are applied as benchmark for performance testing. Precision, recall, and F-score are used as the evaluation criteria. The precision of different methods is shown in Table 2. Their recall and F-score are displayed in Figures 2 and 3. For the English entity relationship extraction data set, the F-score of six models is listed in Table 3 for the scalability evaluation of the proposed model.

Table 2 displays the precision of Bi-LSTM-CRF, Att-Bi-LSTM-CRF, bert-based, and the F-scores of our methods varying from 72.5% to 79.2%. The proposed Bi-MEMM method enjoys the highest precision of 79.2%, while precision values of the other methods are 72.5%, 73.6% and 75.1%. In terms of recall and F-score shown in Figures 2 and 3, it can be concluded that our model performs efficiently with a highest recall of 80.4% and a outstanding F-score of 79.8%, while those of other methods are 71.6% and 72.05% (Bi-LSTM-CRF), 74.3% and 73.94% (Att-Bi-LSTM-CRF), and 76.3% and 75.69% (bert-based).

Bi-MEMM has some features which can overcome the pitiful of traditional methods while dealing with Chinese entity relationship extraction. Firstly, the MEMM model, like the CRF model, has an attractive feature with the convexity of its loss function. The Bi-MEMM model fundamentally solves the label bias problem of the MEMM model and can make full use of context information. It can use complex, overlapping, and nonindependent information for its training and inference. Compared with the CRF model, the performance of feature selection in the Bi-

Table 2: Chinese entity relationship extraction precision.

Methods	Precision
Bi-LSTM-CRF [5]	72.5
Att-Bi-LSTM-CRF [26]	73.6
Bert-based [27]	75.1
Our method	79.2

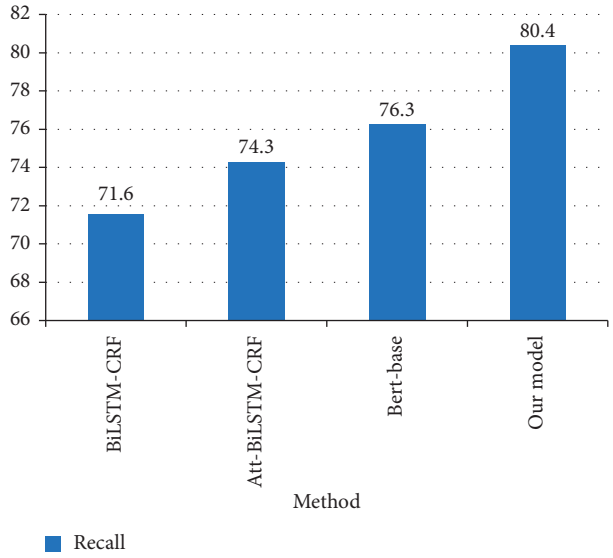


Figure 2: Recall of different methods.

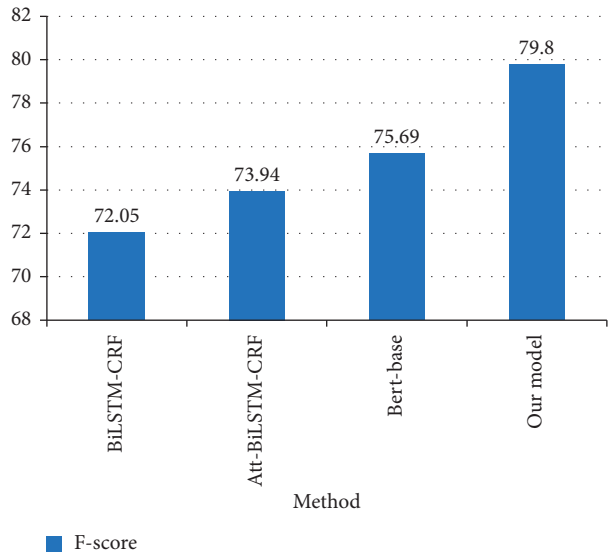


Figure 3: F-score of different methods.

MEMM model is no longer directly determining the level of system performance. Secondly, the entity relationship chain we proposed can efficiently tackle the problems as entity overlap and relationship intersection without the following two shortcomings. The first is error accumulation and entity redundancy caused by the mutual influence of entity recognition and relationship extraction which can lead to the computational complexity; the second is the lack of interaction information caused by ignoring the internal connection and dependency between the entity recognition and relationship extraction.

Table 3: English entity relationship extraction results.

Methods	F-score
BRNN [28]	82.5
SDP-BLSTM [29]	83.7
CNN [30]	77.5
Att-RCNN [31]	83.7
Hybrid Bi-LSTM-Siamese [32]	81.8
Our method	84.60

Table 3 reveals the scalability of our proposed method, which can handle the English entity relationship extraction. Moreover, our method can reach an outstanding F-score of 84.6% which is overall higher than that of the other five methods. The results indicate that the proposed method not only performs well in dealing with Chinese entity relationship extraction but also has a superior scalability while dealing with English.

5. Summary and Future Work

In this paper, a joint extraction model based on joint coding is proposed, and Bi-MEMM is introduced into the joint extraction model and applied to entity relationship extraction tasks. Experiments show that the model performs well in Chinese data sets and has a strong scalability in English data sets. It owns the ability to learn the internal structure of a sentence without considering the complexity of named entities and relationships in the sentence. At the same time, we also notice that the model is still inadequate in dealing with the long-distance constraint of sample sentences, the implicit relation in entities, the reasoning of the same relation such as referential relation, subordination relation, and date writing format problem. Of course, annotation set data is also an important factor that cannot be ignored. We expect that future work could be carried out from the following aspects, such as integrating natural language algorithms (e.g., anaphora resolution into Deep Learning algorithms) and external knowledge bases (e.g., thesaurus, WordNet, HowNet, and knowledge map prior validation) waiting to be introduced into the model. We believe that the introduction of these methods in future modelling will greatly improve their accuracy.

Data Availability

The data used to support the findings of this study are available from the following website list: People's Daily: https://github.com/buppt/ChineseNER; English SemEval 2010 Task 8: https://www.kaggle.com/drtoshi/semeval2010-task-8-dataset.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This paper was partially supported by the National Natural Science Foundation of China (Nos. U1711266, 62076224, and 41925007), Department of Education Neo-Generation Information Technology Innovation Project (Nos. 2018A03006 and 2018A02021), and General Project of Education Humanities and Social Sciences in Hubei Province (No. 18Y38).

References

[1] C. Aone and M. Ramos-Santacruz, "REES: a large-scale relation and event extraction system," in *Proceedings of the Sixth Applied Natural Language Processing Conference*, pp. 76–83, Seattle, WA, USA, April 2000.

[2] J. S. Aitken, "Learning information extraction rules: an inductive logic programming approach," in *Proceedings of the ECAI 2002*, pp. 355–359, Lyon, France, July 2002.

[3] A. Schutz and P. Buitelaar, "Relext: a tool for relation extraction from text in ontology extension," in *Proceedings of the International Semantic Web Conference 2005*, pp. 593–606, Galway, Ireland, November 2005.

[4] B. Rink and S. Harabagiu, "A generative model for unsupervised discovery of relations and argument classes from clinical texts," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 519–528, Edinburgh, Scotland, July 2011.

[5] S. Thattinaphanich and S. Prom-on, "Thai named entity recognition using Bi-LSTM-CRF with word and character representation," in *Proceedings of the 2019 4th International Conference on Information Technology (InCIT)*, pp. 149–154, Bangkok, Thailand, October 2019.

[6] A. Culotta and J. Sorensen, "Dependency tree kernels for relation extraction," in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04) 2004*, pp. 423–429, Barcelona, Spain, July.

[7] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu, "Joint extraction of entities and relations based on a novel tagging scheme," 2017, https://arxiv.org/abs/1706.05075.

[8] B. Deng, X. Fan, and L. Yang, "Entity relation extraction method using semantic pattern," *Jisuanji Gongcheng/Computer Engineering*, vol. 33, no. 10, pp. 212–214, 2007.

[9] R. McDonald, F. Pereira, S. Kulick, S. Winters, Y. Jin, and P. White, "Simple algorithms for complex relation extraction with applications to biomedical IE," in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05) 2005*, pp. 491–498, Ann Arbor, MI, USA, June 2005.

[10] C. Wen, Z.-X. Shi, and Y. Xin, "Chinese non-taxonomic relation extraction based on extended association rule," *Computer Engineering*, vol. 24, 2009.

[11] E. Riloff and R. Jones, "Learning dictionaries for information extraction by multi-level bootstrapping," in *Proceedings of the AAAI/IAAI 1999*, pp. 474–479, Orlando, FL, USA, July 1999.

[12] M. Sabou, M. d'Aquin, and E. Motta, "SCARLET: semantic relation discovery by harvesting online ontologies," in *Proceedings of the European Semantic Web Conference 2008*, pp. 854–858, Karlsruhe, Germany, October 2008.

[13] H. Yu, L. Qian, G. Zhou, and Q. Zhu, "Chinese semantic relation extraction based on unified syntactic and entity semantic tree," *Journal of Chinese Information Processing*, vol. 24, no. 5, pp. 17–23, 2010.

[14] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld, "Knowledge-based weak supervision for information extraction of overlapping relations," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies 2011*, pp. 541–550, Portland, ON, USA, June 2011.

[15] Z. Qin and F. Ye, "Research on reliability of instance and pattern in semi-supervised entity relation extraction," in *Recent Developments in Intelligent Computing, Communication and Devices*, pp. 377–385, Springer, Berlin, Germany, November 2019.

[16] D. Davidov, A. Rappoport, and M. Koppel, "Fully unsupervised discovery of concept-specific relationships by web mining," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics 2007*, pp. 232–239, Prague, Czech Republic, June 2007.

[17] Y. Yan, N. Okazaki, Y. Matsuo, Z. Yang, and M. Ishizuka, "Unsupervised relation extraction by mining wikipedia texts using information from the web," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP 2009*, pp. 1021–1029, Singapore, August 2009.

[18] T.-J. Fu, P.-H. Li, and W.-Y. Ma, "GraphRel: modeling text as relational graphs for joint entity and relation extraction," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics 2019*, pp. 1409–1418, Italy; Venue, January 2019.

[19] Y. Zhang, Z. Guo, and W. Lu, "Attention guided graph convolutional networks for relation extraction," 2019, https://arxiv.org/abs/1906.07510.

[20] Z. He, Z. Zhou, L. Gan, J. Huang, and Y. Zeng, "Chinese entity attributes extraction based on bidirectional LSTM networks," *International Journal of Computational Science and Engineering*, vol. 18, no. 1, pp. 65–71, 2019.

[21] F. Bai and A. Ritter, "Structured minimally supervised learning for neural relation extraction," 2019, https://arxiv.org/abs/1904.00118.

[22] G. Ji, K. Liu, S. He, and J. Zhao, "Distant supervision for relation extraction with sentence-level attention and entity descriptions," in *Proceedings of the AAAI 2017*, Francisco, CA, USA, February 2017.

[23] G. Bekoulis, J. Deleu, T. Demeester, and C. Develder, "Adversarial training for multi-context joint entity and relation extraction," 2018, https://arxiv.org/abs/1808.06876.

[24] H. Xun, Y. Hongliang, and Y. Yang, "A review of relation extraction," *Data Analysis and Knowledge Discovery*, vol. 29, no. 11, pp. 30–39, 2013.

[25] W. Zhang, Y. Feng, F. Meng, D. You, and Q. Liu, "Bridging the gap between training and inference for neural machine translation," 2019, https://arxiv.org/abs/1906.02448.

[26] L. Luo, Z. Yang, P. Yang et al., "An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition," *Bioinformatics*, vol. 34, no. 8, pp. 1381–1388, 2018.

[27] W. Huang, X. Cheng, T. Wang, and W. Chu, "BERT-based multi-head selection for joint entity-relation extraction," in *Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing 2019*, pp. 713–723, Zhengzhou, China, October 2019.

[28] D. Zhang and D. Wang, "Relation classification via recurrent neural network," 2015, https://arxiv.org/abs/1508.01006.

[29] X. Yan, L. Mou, G. Li, Y. Chen, H. Peng, and Z. Jin, "Classifying relations via long short term memory networks along

shortest dependency path," 2015, https://arxiv.org/abs/1508.03720.

[30] Y. Wang, *Entity Relationship Extraction Based on Deep Convolutional Neural Network*, Taiyuan University of Technology, Taiyuan, China, 2017.

[31] X. Zhang, F. Chen, and R. Huang, "A combination of RNN and CNN for attention-based relation classification," *Procedia Computer Science*, vol. 131, pp. 911–917, 2018.

[32] Z. Cui, L. Pan, and S. Liu, "Hybrid bilstm-siamese network for relation extraction," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems 2019*, pp. 1907–1909, Montreal, Canada, May 2019.

[33] B. Srinivasa-Desikan, *Natural Language Processing and Computational Linguistics: A Practical Guide to Text Analysis with Python, Gensim, spaCy, and Keras*, Packt Publishing Ltd, Birmingham, UK, 2018.

[34] B. Wang, R. Hu, and L. Yang, "Constructing the image graph of tang poetry," in *Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing 2019*, pp. 426–434, Dunhuang, China, October 2019.

[35] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford CoreNLP natural language processing toolkit," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations 2014*, pp. 55–60, Baltimore, MD, USA, June 2014.

Hindawi
Complexity
Volume 2021, Article ID 6061939, 10 pages
https://doi.org/10.1155/2021/6061939

Research Article

Vehicle Type Recognition Algorithm Based on Improved Network in Network

Erxi Zhu,[1,2] **Min Xu**,[3,4] **and De Chang Pi**[2]

[1]*College of Internet of Things Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China*
[2]*College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China*
[3]*College of Electronic and Information Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China*
[4]*Jiangsu Key Laboratory of ASIC Design, Wuxi 214153, China*

Correspondence should be addressed to Erxi Zhu; erxi666@163.com

Received 7 July 2020; Revised 19 November 2020; Accepted 22 December 2020; Published 5 January 2021

Academic Editor: Jia Wu

Vehicle type recognition algorithms are broadly used in intelligent transportation, but the accuracy of the algorithms cannot meet the requirements of production application. For the high efficiency of the multilayer perceptive layer of Network in Network (NIN), the nonlinear features of local receptive field images can be extracted. Global average pooling (GAP) can avoid the network from overfitting, and small convolution kernel can decrease the dimensionality of the feature map, as well as downregulate the number of model training parameters. On that basis, the residual error is adopted to build a novel NIN model by altering the size and layout of the original convolution kernel of NIN. The feasibility of the algorithm is verified based on the Stanford Cars dataset. By properly setting weights and learning rates, the accuracy of the NIN model for vehicle type recognition reaches 97.2%.

1. Introduction

Intelligent transportation [1] refers to a research hotspot in existing society, and vehicle type recognition [2] underpins and critically impacts intelligent transportation studies. The existing algorithms of vehicle type recognition are primarily classified as manual feature descriptions, 3D model, and artificial intelligence algorithms. At the early phase, the manual feature descriptions (e.g., SIFT [3] and HOG [4]) are adopted to extract vehicle features; subsequently, the algorithms (e.g., SVM and decision tree) are combined for classification. Since feature extraction and data reconstruction are difficult to achieve, Hsieh et al. [5] employed HOG and symmetric SURF descriptor to extract the vehicle features of mesh generation. Besides, Liao et al. [6] conducted the appearance and semantic segmentation of vehicle parts to recognize vehicle types. Moreover, Biglari et al. [7] exploited the overall appearance of the vehicles and the feature differences of various components to train the SVM classifier. The mentioned algorithms are easy to affect by environmental factors (e.g., light and background), so their recognition accuracy is relatively low. As impacted by the random variation in the shooting angle of vehicle images, the 3D model-based vehicle type recognition method was developed at the right moment. The 3D model can reflect spatial relationships between local features and the whole vehicle. Existing studies [8, 9] effectively performed the 3D modeling and feature extraction of vehicles. Artificial intelligence introduced a novel impetus into vehicle type recognition, and the features of the vehicle can be automatically extracted. Dong et al. [10] adopted the sparse Laplace filter and a semisupervised convolution neural network to extract vehicle features and classify vehicles. Studies [11–14] employed different methods or optimized the existing neural network to conduct the vehicle type recognition, and its effect was significantly improved; however, for the similar vehicle recognition exhibiting a remarkably small feature gap (e.g., Volkswagen's front face is nearly identical), the room for improvement of classification accuracy is limited.

In view of the low accuracy of vehicle type recognition, we propose an improved NIN for vehicle type recognition and get

high recognition accuracy. In fact, the breakthrough point of vehicle type recognition refers to the efficient extraction of nonlinear features of vehicles. NIN [15] exhibits a complex multilayer perceptron (MLPConv) with a micronetwork structure and is capable of efficiently and automatically extracting local nonlinear features of images. The present study fully exploits the following features of the NIN model and uses its 1×1 convolution kernel to conduct the dimensionality reduction of the feature map and downregulate the number of network parameters. The global average pooling layer (GAP) is adopted to effectively combine the features and prevent the whole network from falling into the overfitting state. The improvement measures are as follows: the original large convolution kernel of NIN is changed into a small convolution kernel, which increases the depth of convolution neural network and improves the performance of the network. In order to avoid the gradient loss problem caused by the increase of depth, residual measures are arranged on the structure to solve the network degradation problem. The improved NIN has high classification effects, and its classification accuracy is better than VGG and GoogLeNet in vehicle type recognition. By the verification based on the Stanford Cars dataset and the reasonable weight and learning rate setting, the vehicle type recognition accuracy of the improved NIN reaches over 97.2%.

2. Related Works

The 1×1 small convolution kernel, GAP, micronetwork structure, and other measures proposed by NIN underpin the follow-up deep convolutional neural network (CNN). CNN [16] automatically extracts image features; thus, the complex feature extraction and data reconstruction process of conventional recognition algorithms can be avoided. AlexNet [17], VGGNet [18–21], GoogLeNet [22, 23], ResNet [24–27], and other networks can be adopted for vehicle type recognition, whereas for the limitations of sample quality and quantity as well as the defects of network feature extraction and classification performance, vehicle recognition exhibits relatively low accuracy.

Most networks are only capable of extracting linear features on the images, landing the classification algorithm in confusion since the linear features are basically consistent (Figure 1(a) and (b)). For classification, only the overall information built by linear features can be classified (Figure 1(c)).

In Figure 1, the linear features denoted by (a) and (b) are consistent, which are both a line segment and a part of an object without any difference. However, given the overall information, the information represented by (c) is completely inconsistent. Thus, a question is raised of how to extract this nonlinear feature effectively. This question is determined by the micronetwork [28, 29] structure embedded in NIN, i.e., a full connection layer consisting of two layers of convolution. In the neural network, two-layer fully connected hidden neurons are capable of approximating arbitrary curves.

2.1. "Micronetwork" Structure. In 2013, the proposal of NIN modified the original idea of network structure, and the multilayer perceptron was built by replacing the conventional linear perceptron with the embedded "micronetwork"; as a result, the efficiency of nonlinear feature extraction of local sensing field of images was significantly enhanced.

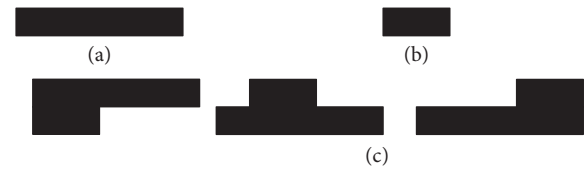


Figure 1: Schematic diagram of linear features and linear feature combinations.

In NIN, "micronetwork" refers to a general nonlinear function approximator. The difference between MLPConv of NIN and linear perceptron of CNN is the method of image feature extraction. MLPConv consists of several fully connected nonlinear activation functions, shared by all local receptive fields. Moreover, by sliding on the input, the feature map is generated and then outputted to the next layer. MLPConv can combine different feature maps, so the network can extract complex and useful nonlinear image features. Furthermore, the overall structure of NIN can be superposed by multiple MLPConv.

There are two reasons why NIN selects multilayer perceptron: (1) MLPConv fits the structure of the convolutional neural network and (2) MLPConv can act as a deep model, complying with the spirit of feature reuse [22]. The feature map of MLPConv is calculated:

$$\begin{aligned} f^1_{i,j,k_1} &= \max\left(\omega^{1\ T}_{k_1} x_{i,j} + b_{k_1}, 0\right), \\ &\vdots \\ f^n_{i,j,k_n} &= \max\left(\omega^{n\ T}_{k_n} f^{n-1}_{i,j} + b_{k_n}, 0\right), \end{aligned} \tag{1}$$

where n denotes the number of layers of the multilayer perceptron; (i, j) represents the pixel index in the feature map; $x_{i,j}$ indicates the input block centred on the position (i, j); k is the channel index of the feature map; and b_{k_1} is the bias. ReLU acts as the activation function in MLPConv.

2.2. Global Average Pooling Layer. In the classification, GAP [30, 31] remedies the defect of the fully connected layer. At the early phase, the feature map of the final convolutional layer is vectorized and passed into the fully connected layer; subsequently, it is inputted to the SoftMax layer [32–34]. Since the fully connected layer is easy to overfit, the whole network exhibits a reduced generalization ability, and the subsequent network conducts a dropout [24] operation on the fully connected layer, thereby preventing overfitting significantly. However, GAP is adopted by NIN to set the last MLPConv feature map to pertain to the corresponding classification category, which can more effectively fit the convolution structure. There are no parameters to be optimized in the operation, thereby avoiding overfitting. The regularization effect of GAP is more significant than dropout.

2.3. 1×1 *Convolution Kernel.* The 1×1 convolution was initially proposed by NIN to make the network exhibit significantly high network performance. By 1×1 convolution computation, MLPConv reduces the dimension of the channel parameter pool of convolutional kernel, as well as downregulating the number of parameters. The main functions of 1×1 convolution are as follows:

(1) Dimensionality reduction: for instance, if an 500×500 image with a depth of 100 is generated with 1×1 convolution on 20 filters, the size of the result is $500 \times 500 \times 20$.

(2) The nonlinear expression ability is enhanced. After the convolutional layer passes through the excitation layer, the 1×1 convolution introduces nonlinear excitation to the learning representation of the previous layer to enhance the expression ability of the network.

(3) The model depth is increased. Accordingly, the number of the network model parameters can be reduced, the depth of the network layer can increase, and the representational capacity of the model can be enhanced to some extent.

Figure 2 illustrates the NIN structure of 4 MLPConv and 1 GAP. Subsampling layers can be added between MLPConv layers, and the number of layers of the "micronetwork" can be altered for specific tasks. First, taking the first MLPConv as an example, the input image is $224 \times 224 \times 3$, 224 represents the pixel of the input image, and 3 denotes the channel of the image. Later, the convolution filter is adopted to slide on the input image and calculate the inner product. The size of the convolution filter adopts $11 \times 11 \times 3$, i.e., the length and width are both 11, and the depth is 3. In the first layer of MLPConv, 96 convolution filters are adopted. The embedded "micronetwork" refers to a fully connected neural network with a two-layer convolutional kernel, performing nonlinear feature extraction. The number of neurons in each layer reaches 96. Besides, Figure 2 presents one of the models compared in subsequent experiments, and the specific setting of parameters is presented in the figure.

In the present study, the nonlinear feature extraction capacity of NIN is exploited to extract the features of vehicles in the image (e.g., texture and topology structure) to enhance the efficiency of the vehicle type recognition. On that basis, by increasing the size, quantity, and layout of the convolutional kernel in NIN, as well as the network performance and convergence speed, the training of NIN for vehicle sample data is conducted efficiently, and the vehicle recognition accuracy is enhanced. Subsequently, the residual thought is adopted to solve gradient dissipation that is attributed to the rising number of network layers.

3. Optimized NIN

At present, network performance can be enhanced primarily by two measures. One is to increase the width or depth of the network. For instance, VGG enhances network performance by increasing network depth. The other refers to optimizing the network input sample data (e.g., increasing the sample number, strengthening the texture of the sample, or transforming the shape of the sample image (inversion and distortion) to enhance the network performance). For the deepened or widened network, its defects gradually appear, the gradient disappears, the number of parameters is huge, and the extracted features tend to be invalid in the network transmission. In the present study, NIN is optimized by the following two means.

3.1. Use of Small Convolution Kernel. The small convolution kernel increases the network depth and improves the network performance, as well as significantly downregulates the number of network parameters. In numerous networks, the convolution kernel with a size of 3×3 and 5×5 has been extensively used, and 3×3 refers to the smallest size that can capture 8 neighbourhood information of pixels.

The small convolution kernels are stacked to replace the large convolution kernels, and the size of the receptive field remains unchanged. Multiple 3×3 convolution kernels exhibit more nonlinearities (more layers of nonlinear functions) than the convolution layer of a large convolution kernel. Moreover, multiple 3×3 convolutional layers have fewer parameters than a large convolution kernel. If the input and output feature maps of the convolutional layer are assumed to have an identical size to C, the number of parameters of the three convolutional layers is $3 \times (3 \times 3 \times C \times C) = 27C^2$. The parameter of one 7×7 convolutional layer is $49C^2$. Thus, the small convolution kernel significantly reduces the number of network parameters.

At the beginning of AlexNet and NIN training, a large convolution kernel is employed for calculation, and the classification accuracy is not significantly enhanced. Even though NIN employs a micronetwork as a local nonlinear feature collector, it only increases the convergence speed of the model. On the whole, the convolution kernel of VGG uses 3×3 convolution kernel, and GoogLeNet contains 3×3, 5×5, and 1×1; the classification effect of VGG and GoogLeNet models is larger than that of the former two. Indeed, this is also attributed to the deepening of the number of network layers. The function of 1×1 convolution kernel suggested that it exhibits the function of raising and reducing dimension and can downregulate the number of network parameters in Section 2.

An experiment is performed to verify the influence of small convolution on the model classification. MINST dataset is employed in the experiment, and the network structure is adopted (Figure 3). The experiment is split into two groups to verify the effect of 7×7, 5×5, and 3×3 convolution kernels on the network performance, respectively. The statistics is summarized to the iteration times under the accuracy of the four models reaching over 0.6, 0.7, 0.8, and 0.9 initially, as well as the iteration times in the presence of maximum accuracy as well as the maximum accuracy and time consumed initially. Each model experiment is repeated 50 times, and the average number of statistical iterations is listed in Table 1.

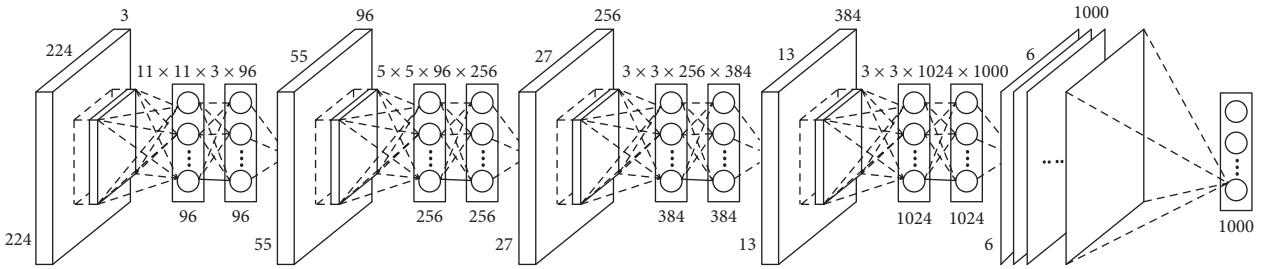


Figure 2: Structure and specific parameter settings of NIN.

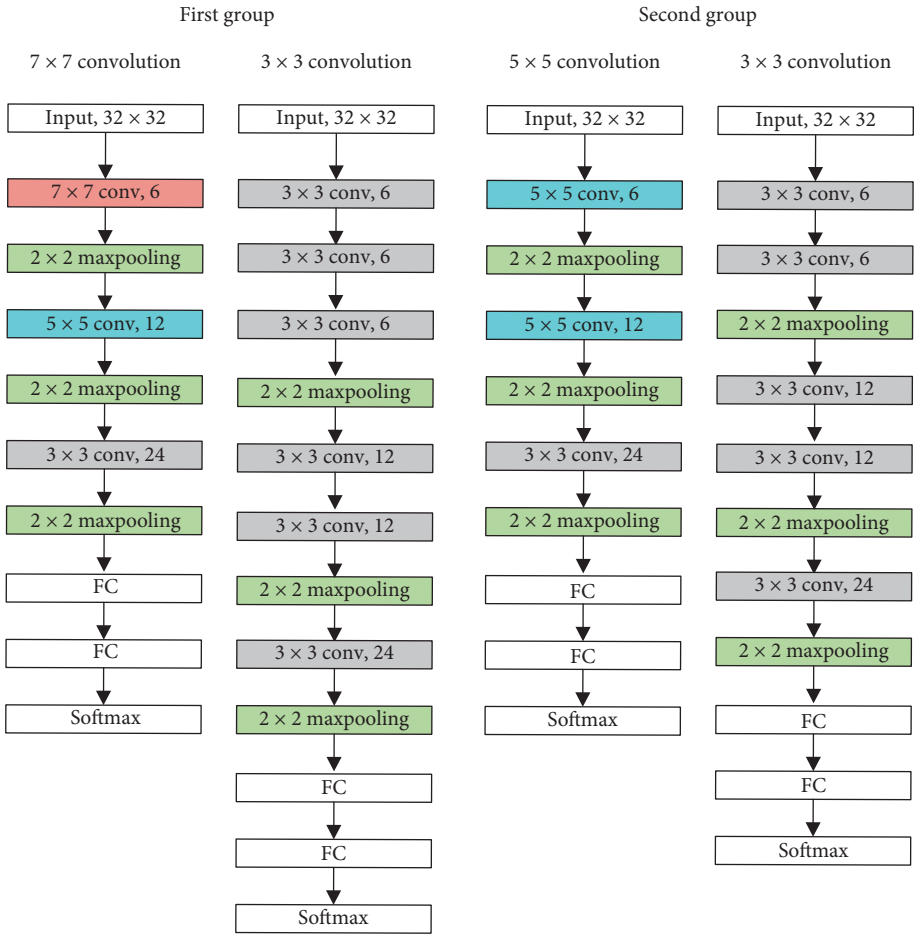


Figure 3: Network structure of small convolutional kernel experiment.

Table 2 presents that the small convolution kernel enhances the extraction performance of local receptive field features of the network and increases the classification accuracy of the model. Three 3×3 convolution kernels are equivalent to a 7×7 convolution kernel, and two 3×3 convolution kernels are equated with a 5×5 convolution kernel. Under the receptive field of the identical convolution kernel, it is easy to find by comparison that the recognition efficiency of the convolution kernel falls to the maximum. In all effective intervals, the average number of experimental iterations of 5×5 convolution kernel is smaller than that of 3×3 and 7×7 convolution kernels. 3×3 convolution kernel exhibits the highest accuracy, whereas the accuracy of 5×5 convolution kernel is relatively low; however, the convolution kernel exhibits significantly low accuracy. Accordingly, in general, 3×3 convolution kernel has the maximum recognition efficiency and the fastest rise in accuracy; that is, 3×3 convolution kernel exhibits a better performance to extract local features of images.

To obtain the vehicle type recognition accuracy, the NIN structure is optimized. The size, quantity, and layout of the convolution kernel of the NIN structure in Section 2 are tuned in accordance with the advantages of the small convolution kernel to extract local features of the image and downregulate

Table 1: Experimental data of small convolution kernel.

Convolution form	Exceeding the average number of iterations of accuracy for the first time					Maximum accuracy rate	Time (h)
	0.6	0.7	0.8	0.9	Maximum		
(1) 7*7	100.9	157.1	230.6	320.1	832.2	0.845	1.261
(1) 3*3	88.4	148	207.6	293.8	799.3	0.965	1.469
(2) 5*5	75.3	114.5	175.2	252.6	821.6	0.903	1.328
(2) 3*3	74.4	110.0	172.6	249.8	819.3	0.914	1.409

Table 2: Comparison of experimental results on the Stanford Cars dataset.

Network name	1		2		3		4		5	
	Accuracy rate (%)	Iterations	Accuracy rate (%)	Iterations	Accuracy rate (%)	Iterations	Accuracy rate (%)	Iterations	Accuracy rate (%)	Iterations
NIN	80.2	5066	88.3	5628	90.4	5822	91.6	5923	91.2	6022
VGG19	83.1	5732	89.1	5913	91.5	6134	92.4	6417	92.7	6982
GoogLeNet	85.1	5522	90.3	5817	92.6	6025	93.0	6120	94.3	6216
New NIN	84.2	4909	90.5	5423	95.5	5781	96.2	5883	97.2	5989

the number of computational parameters of the network. Figure 4 suggests that the 11×11 convolution kernel of the first layer is converted into $4_{3\times3}$ convolution kernels.

3.2. Use of Residual Blocks. Since AlexNet, the depth of the most advanced CNN architecture has been increasing, whereas the depth of the network cannot increase by simply stacking layers. The mentioned finding is because the gradient backpropagates to the previous layer, and repeated multiplication may make the gradient infinitesimal and the gradient disappear; the deep network is difficult to train, and the network performance tends to be saturated, or even drops rapidly. To address this problem, He Kaiming et al. proposed the residual network ResNet; in 2015, the proposed network won the first prize in the challenge competition of ImageNet image recognition and has deeply inspired the design of the later deep neural network.

He Kaiming considered that the training errors produced by stacking identity maps on the deep network should not be higher than those attributed to shallow networks. According to Figure 5, the residual block can achieve the mentioned condition, and the input can be spread by cross-layer data line forward faster. In fact, ResNet is not the first model exploiting fast connection. Highway networks [35] and long and short-term memory network [36] units employ different gate structures to conduct fast connection.

ResNet (Figure 6) continues to use the design of all 3×3 convolution layer of VGG. First, there are two 3×3 convolutional layers with an identical number of output channels in the residual block. Each convolutional layer is followed by a batch normalization layer and ReLU activation function. Subsequently, the input is directly introduced to the front of the final ReLU activation function by skipping the two convolutional operations. In the mentioned design, the output and input of the two convolutional layers should exhibit the identical shape, and then they should be added. To alter the number of channels, an additional 1×1 convolutional layer should be introduced to transform the input into the required shape, and then an addition operation is required.

As impacted by small convolution kernel and residual concept, the NIN is further optimized, and the convolution kernel in NIN is replaced by 3×3 convolution kernel to conduct the rapid convergence and training of the network. The residual measurement is performed to build data lines between the front and back layers of the network, so the feature map can be efficiently transmitted to the front convolutional layer, thereby eliminating the effect of gradient accumulation and decreasing and avoiding gradient disappearance. Given the setting requirements of ResNet, the optimized NIN structure is illustrated in Figure 6.

4. Implementation of Optimized NIN

The optimized NIN uses 3×3 convolution kernel and 1×1 convolution kernel [37, 38]. 3×3 convolution kernel is used to increase network depth and improve network performance. 1×1 convolution kernel is used to enhance the extraction ability of nonlinear features of the network. In the optimized NIN structure, GAP is used as a classifier instead of full connection layer and to improve the generalization ability of the network and avoid overfitting of the network. In order to avoid the loss of gradient caused by the increase of network depth, residual measures are arranged between consecutive multiple 3×3 convolution layers on the optimized NIN to avoid network degradation. The partial source code of optimized NIN is as follows: (Algorithm 1)

5. Results and Discussion

The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed sections. The representative Stanford Cars dataset is adopted in the experiment. The scene with the images located varies with different postures [39] and unfixed resolutions. Accordingly, the vehicle type recognition of this dataset is more challenging. The Stanford Cars dataset

```
Input:
input_shape: Input shape of network, default as (224,224,3)
nclass: Numbers of class (output shape of network), default as 1000
Output: Optimized NIN model
The optimized NIN model is established according to the following steps:
Step 1: Build two residual blocks including 384 convolution kernels
            Build two 1 × 1 convolution layers
            Build Max pool layer
Step 2: Build two residual blocks including 384 convolution kernels
            Build two 1 × 1 convolution layers
            Build Max pool layer
Step 3: Build residual block including 384 convolution kernels
            Build two 1 × 1 convolution layers
            Build Max pool layer
Step 4: Build residual block including 1024 convolution kernels
            Build two 1 × 1 convolution layers
Step 5: Build GAP layers
            return model
```

Algorithm 1: Partial source code of optimized NIN.

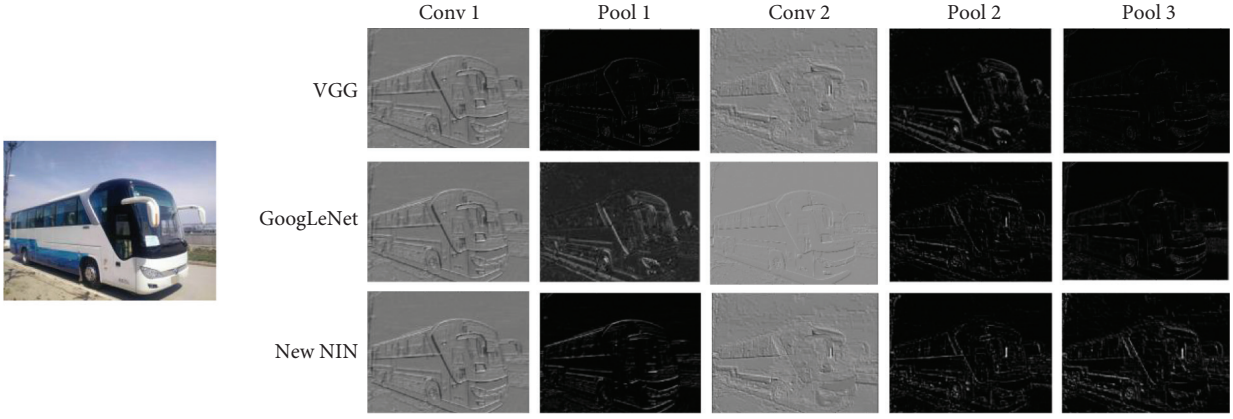


Figure 4: Visualization of activation values in the middle layer of three types of convolutional neural network models.

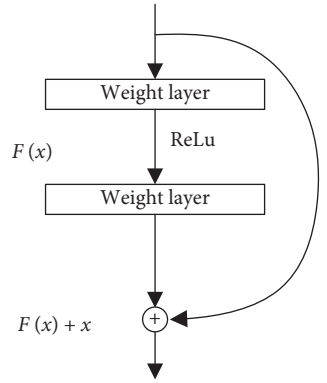


Figure 5: Schematic diagram of the residual block.

consists of 196 vehicle types, containing 16,185 images overall. The dataset labels consist of the vehicle types and the location of the vehicles in the image. The hardware environment of the experiment is presented: CPU type is Xeon W; memory type is DDR4 128GB; graphics card is NVIDIA RTX 2080Ti, and video memory size is 11GB. All the experimental networks are achieved by GPU built by Anaconda 3 + Tensorflow 2.0 + Spyder + Python 3.7 in Windows 10.

To determine the performance of optimized NIN on vehicle type feature extraction, VGG19 (layer 19), GoogLeNet Inception V1 (layer 22), NIN (layer 12), and

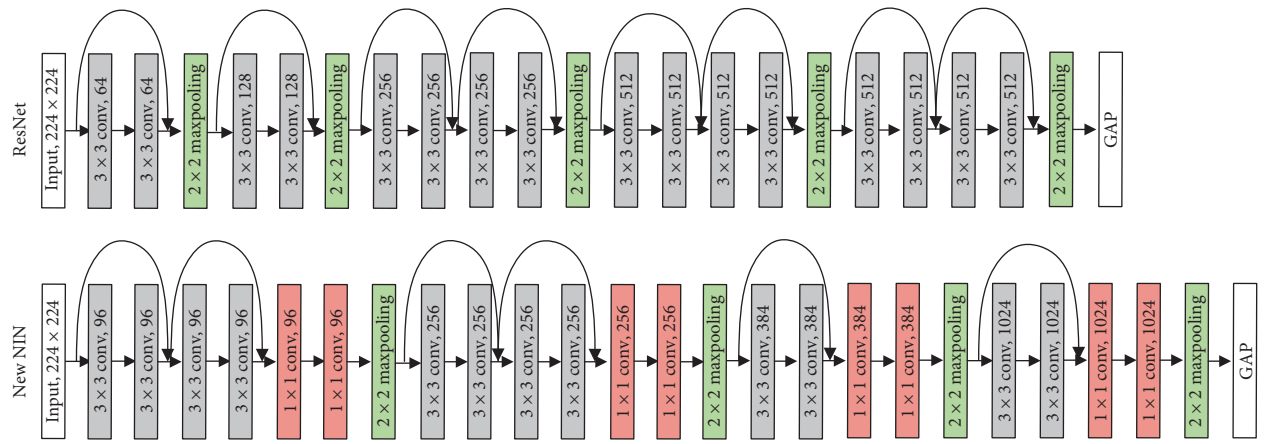


Figure 6: ResNet network structure and optimized NIN structure.

optimized NIN (layer 20) act as the comparison network models. GAP + SoftMax is employed for all the mentioned network model classifiers, and all network training employs data input dimensions. The preprocessing of the dataset, the splitting of the training set, and the verification set comply with literature [40]: the image size of the dataset is normalized to 256×256, 4 corners and the centre part are cut to generate 5 images with a size of 224×224, and the mirror operation is performed to generate 10 training images on the whole, from which the mean value of the training set image is subtracted to obtain the training input data. In the present study, appropriate weights and learning rates are manually set to achieve initialization. The training process starts from the initial weight and learning rate and continues till the accuracy of the training set stops enhancing, and then the learning rate reduces to one-tenth of the original. This process is repeated five times. The weight of the model is updated with the stochastic gradient descent method, and the initial learning rate is 0.01.

5.1. Vehicle Type Recognition Performance. After repeated training of several models, the classification accuracy rate and the number of iterations reached initially are determined from the Stanford Cars sample data, as listed in Table 2.

The optimized NIN has the original MLPConv of NIN. The nonlinear features of the image can be approximated through "micronetwork" structure, so the optimized NIN has fast convergence. By replacing the large convolution kernel of the original NIN with the small convolution kernel, the optimized NIN has deeper layers than the original NIN. The computational effect of multiple 3×3 convolution kernels is equivalent to that of a 5×5 convolution kernel. Using this conversion, all the large convolution kernels of the original NIN are replaced by 3×3 small convolution kernels, which increases the convolution layers of the NIN and enhances the network performance. The residuals are deployed on the NIN structure to avoid the loss of gradient and restrain the degradation of network performance. It can be found from Table 2 that the number of iterations of NIN in each iteration process is less than that of VGG and

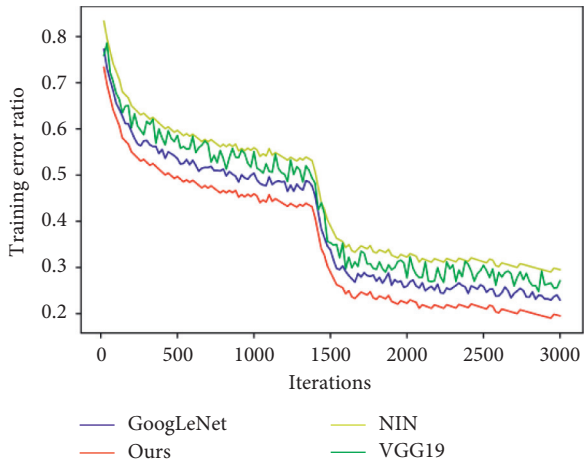


Figure 7: Training error curve.

GoogLeNet, which indicates that the convergence speed of NIN is quicker than that of VGG and GoogLeNet. However, at the end of the experiment, the recognition accuracy of NIN did not exceed that of VGG and GoogLeNet, even if the NIN trained too many iterations. However, the optimized NIN keeps good convergence because the "micronetwork" structure can extract the nonlinear features of the automobile image. In addition, the optimized NIN solves the problem of gradient weakening in the calculation process by the residual layout and strengthens the feature map for subsequent calculation. Therefore, the optimized NIN model outperforms VGG and GoogLeNet in accuracy and convergence speed, and the final vehicle type recognition accuracy reaches 97.2%.

5.2. Vehicle Feature Extraction Capability. VGG19 and GoogLeNet only consist of linear perception layer and only extract linear features [41–45] of vehicles, while NIN and optimized NIN contain multilayer perception layer, which can capture nonlinear features of vehicles. Figure 4 draws the comparison of feature maps of feature extraction of vehicle images after training of several network models.

In Figure 4, Column Conv1 presents the effect of feature map extraction of the three networks after the first convolution kernel operation, column Pool1 refers to the effect of the first pooling layer processing, and column Conv2 represents the sixth-layer convolution calculation results of VGG19 and the third inception structure processing result of GoogLeNet, as well as the second MLPConv processing result of NIN. As revealed from the figure, the ability of the optimized NIN model to extract feature map reaches over those of VGG and GoogLeNet.

5.3. Convergence Effect of Optimized NIN. The experimental data of NIN, VGG19, GoogLeNet, and the optimized NIN in the first 3000 iterations of the third experiment are intercepted, and the training error curves of the sample data of the four networks are plotted (Figure 7).

Figure 7 suggests that the recognition training error of the optimized NIN in the training process is significantly lower than that of the other three networks. In the vicinity of 1300 iterations, the training error of the optimized NIN model did not continue to decrease. We reduce the learning rate of the models participating in the comparison to one-tenth of the original. Each model continued to learn according to the new learning rate, and the training error had a cliff drop in this case, which improves the training speed. In the 3000th iteration, it drops to 19.6%, while the error rate of NIN, VGG19, and GoogLeNet reduces to 31.2%, 28.9%, and 24.6%, respectively. This also indicates that the optimized NIN exhibits good convergence and accelerates the training speed of vehicle license plate recognition.

6. Conclusions

In the present study, the structure and vital components of NIN are analysed, and it is verified that the NIN embedded micronetwork can efficiently extract the nonlinear features of vehicle images, and GAP avoids the overfitting of models and can regularize operation; besides, 1×1 small convolution conducts the dimensionality reduction of feature maps, downregulating the number of model parameters. Based on the NIN, a novel vehicle type recognition algorithm is built by changing the size and layout of the convolution kernel and using residual thought of NIN. Subsequently, it is verified in the Stanford Cars dataset, and the result reveals that the algorithm exhibits a better vehicle type recognition performance and higher recognition accuracy that reaches 97.2%. However, the optimized NIN also has shortcomings. First, in the same local receptive field, the large convolution kernel can be replaced by the small convolution kernel. Although the small convolution kernel operation reduces the number of variables compared with the large convolution kernel operation, the training time is greatly improved, and the efficiency is reduced. Second, the strategy of optimizing NIN is to deepen the network level. To some extent, the application of residual can solve the problem of gradient vanishing and restrain the degradation of network performance. Whether this network performance improvement method can support the further increase of network depth remains to be studied, which also points out the direction for our future work.

Data Availability

The authors used the vehicle dataset provided by Stanford University to verify the improved model. The Cars dataset contains 16,185 images of 196 classes of cars. The data are split into 8,144 training images and 8,041 testing images, where each class has been split roughly in a 50-50 split. Classes are typically at the level of make, model, and year, for example, 2012 Tesla Model S or 2012 BMW M3 coupe; visit http://ai.stanford.edu/~jkrause/cars/car_dataset.html.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the "Geometry Problem Geometry" project (the National Natural Science Foundation of China (NSFC), 61073086). Some of the authors of this publication are also working on these related projects: (1) Higher Vocational Education Teaching Fusion Production Integration Platform Construction Projects of Jiangsu Province under Grant no. 2019(26), (2) Natural Science Fund of Jiangsu Province under Grant no. BK20131097, (3) "Qin Lan Project" Teaching Team in Colleges and Universities of Jiangsu Province under Grant no. 2017(15), and (4) High Level of Jiangsu Province Key Construction Project funding under Grant no. 2017(17).

References

[1] M. Barth and J. J. Sanchez-Medina, "Guest editorial special issue: the 21st IEEE international conference on intelligent transportation systems (ITSC 2018)," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3929-3930, 2020.

[2] Y. Xiang, Y. Fu, and H. Huang, "Global topology constraint network for fine-grained vehicle recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 7, pp. 2918–2929, 2020.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886–893, IEEE, San Diego, CA, USA, June 2005.

[5] J.-W. Hsieh, L.-C. Chen, and D.-Y. Chen, "Symmetrical surf and its applications to vehicle detection and vehicle make and model recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 6–20, 2014.

[6] L. Liao, R. Hu, J. Xiao, Q. Wang, J. Xiao, and J. Chen, "Exploiting effects of parts in fine-grained categorization of vehicles," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 745–749, IEEE, Quebec City, Canada, September 2015.

[7] M. Biglari, A. Soleimani, and H. Hassanpour, "Part-based recognition of vehicle make and model," *IET Image Processing*, vol. 11, no. 7, pp. 483–491, 2017.

[8] Y. L. Lin, V. I. Morariu, W. Hsu, and L. S. Davis, "Jointly optimizing 3D model fitting and fine-grained classification," in *Proceedings of the European Conference on Computer Vision*, pp. 466–480, Springer, Zurich, Switzerland, September 2014.

[9] J. Krause, M. Stark, J. Deng, and F.-F. Li, "3D object representations for fine-grained categorization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 554–561, IEEE, Sydney, NSW, Australia, December 2013.

[10] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.

[11] J. Sochor, A. Herout, and J. Havel, "Boxcars: 3D boxes as cnn input for improved fine-grained vehicle recognition,," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3006–3015, IEEE, Las Vegas, NV, USA, June 2016.

[12] Y. B. Gao and H. J. Lee, "Local tiled deep networks for recognition of vehicle make and model," *Sensors*, vol. 16, no. 2, p. 226, 2016.

[13] S. Yu, Y. Wu, W. Li et al., "A Model for fine-grained vehicle classification based on deep learning," *Neurocomputing*, vol. 257, pp. 97–103, 2017.

[14] B. Hu, J.-H. Lai, and C.-C. Guo, "Location-aware fine-grained vehicle type recognition using multi-task deep networks," *Neurocomputing*, vol. 243, pp. 60–68, 2017.

[15] M. Lin, Q. Chen, and S. Yan, "Network in network," *Computer Science*, arXiv: 1312.4400, 2013.

[16] Y. Hou, L. Zhou, S. Jia, and X. Lun, "A novel approach of decoding eeg four-class motor imagery tasks via scout esi and cnn," *Journal of Neural Engineering*, vol. 17, no. 1, pp. 1–15, 2020.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097–1105, 2012.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.

[19] X. Dai, H. Yin, and N. K. Jha, "Nest: a neural network synthesis tool based on a grow-and-prune paradigm," *IEEE Transactions on Computers*, vol. 68, no. 10, pp. 1487–1497, 2019.

[20] F. Zhang, Y. Liu, Y. Zhou, Q. Yin, and H.-C. Li, "A lossless lightweight CNN design for sar target recognition," *Remote Sensing Letters*, vol. 11, no. 5, pp. 485–494, 2020.

[21] B. Ibrokhimov, C. Hur, and S. Kang, "Effective node selection technique towards sparse learning," *Applied Intelligence*, vol. 50, no. 10, pp. 3239–3251, 2020.

[22] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[23] X. Jin, L. Wu, X. Li et al., "Ilgnet: inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation," *IET Computer Vision*, vol. 13, no. 2, pp. 206–212, 2019.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition,," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.

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

[26] B. Liu, Q. Liu, Z. Zhu, T. Zhang, and Y. Yang, "Msst-resnet: deep multi-scale spatiotemporal features for robust visual object tracking," *Knowledge-Based Systems*, vol. 164, no. 15, pp. 235–252, 2019.

[27] Y. Jiang, Y. Li, and H. Zhang, "Hyperspectral image classification based on 3-D separable resnet and transfer learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1949–1953, 2019.

[28] W. Shao, D. Pi, and Z. Shao, "A Pareto-based estimation of distribution algorithm for solving multiobjective distributed no-wait flow-shop scheduling problem with sequence-dependent setup time," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1344–1360, 2019.

[29] H. Alaeddine and M. Jihene, "Deep network in network," *Neural Computing and Applications*, vol. 2020, Article ID 05008-0, 13 pages, 2020.

[30] X. Zhang and X. Zhang, "Global learnable pooling with enhancing distinctive feature for image classification," *IEEE Access*, vol. 8, pp. 98539–98547, 2020.

[31] W. Gong, H. Chen, Z. Zhang, M. Zhang, and H. Gao, "A data-driven-based fault diagnosis approach for electrical power dc-dc inverter by using modified convolutional neural network with global average pooling and 2-D feature image," *IEEE Access*, vol. 8, pp. 73677–73697, 2020.

[32] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[33] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio et al., "Maxout networks," arXiv: 1302.4389, 2013.

[34] M. D. Zeiler and F. Rob, "Stochastic pooling for regularization of deep convolutional neural networks," arXiv:1301.3557, 2013.

[35] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," *Computer Science*, arXiv:1507.06228, 2015.

[36] E. Zhu and D. Pi, "Photovoltaic generation prediction of CCIPCA combined with LSTM," *Complexity*, vol. 2020, Article ID 1929372, 11 pages, 2020.

[37] W. U. Min, "Application of fuzzy neural network in network fault diagnosis," *Computer Knowledge and Technology*, vol. 14, 2019.

[38] P. Shamsolmoali, M. Zareapoor, and J. Yang, "Convolutional neural network in network (cnnin): hyperspectral image classification and dimensionality reduction," *IET Image Processing*, vol. 13, no. 2, pp. 246–253, 2019.

[39] E. Zhu, M. Xu, and D. Pi, "A novel robust principal component analysis algorithm of nonconvex rank approximation," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9356935, 17 pages, 2020.

[40] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," arXiv:1207.0580, 2012.

[41] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and data engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[42] J. Wu, S. Pan, X. Zhu, and Z Cai, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.

[43] C. Dai, D. Pi, J. Wu, L. Cui, and B. Johnson, "CenEEGs," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 2, pp. 1–25, 2020.

[44] C. Dai, J. Wu, D. Pi et al., "Brain EEG time-series clustering using maximum-weight clique," *IEEE Transactions on Cybernetics*, vol. 2020, pp. 1–15, 2020.

[45] J. Zhu, H. Shi, B. Song, S. Tan, and Y. Tao, "Deep neural network based recursive feature learning for nonlinear dynamic process monitoring," *The Canadian Journal of Chemical Engineering*, vol. 98, no. 4, pp. 919–933, 2020.

Hindawi
Complexity
Volume 2020, Article ID 8715619, 14 pages
https://doi.org/10.1155/2020/8715619

Research Article

Novel Node Centrality-Based Efficient Empirical Robustness Assessment for Directed Network

Xiaolong Deng,[1] **Hao Ding**,[1] **Yong Chen**,[2] **Cai Chen**,[3] **and Tiejun Lv**[1]

[1]*Key Lab of Trustworthy Distributed Computing and Service of Education Ministry, Beijing University of Post and Telecommunication, Beijing 100876, China*
[2]*North Automatic Control Technology Institute, Taiyuan, China*
[3]*China Academy of Information and Communications Technology (CAICT), Beijing 100037, China*

Correspondence should be addressed to Xiaolong Deng; shannondeng@bupt.edu.cn

Received 5 August 2020; Revised 11 October 2020; Accepted 28 October 2020; Published 21 November 2020

Academic Editor: Jia Wu

In recent years, while extensive researches on various networks properties have been proposed and accomplished, little has been proposed and done on network robustness and node vulnerability assessment under cascades in directed large-scale online community networks. In essential, an online directed social network is a group-centered and information spread-dominated online platform which is very different from the traditional undirected social network. Some further research studies have indicated that the online social network has high robustness to random removals of nodes but fails to the intentional attacks, particularly to those attacks based on node betweenness or node directed coefficient. To explore on the robustness of directed social network, in this article, we have proposed two novel node centralities of *ITG* (information transfer gain-based probability clustering coefficient) and $IM_p(v)$ (directed path-based node importance centrality). These two new centrality models are designed to capture this cascading effect in directed online social networks. Furthermore, we also propose a new and highly efficient computing method based on iterations for $IM_p(v)$. Then, with the abundant experiments on the synthetic signed network and real-life networks derived from directed online social media and directed human mobile phone calling network, it has been proved that our *ITG* and $IM_p(v)$ based on directed social network robustness and node vulnerability assessment method is more accurate, efficient, and faster than several traditional centrality methods such as degree and betweenness. And we also have proposed the solid reasoning and proof process of iteration times k in computation of $IM_p(v)$. To the best knowledge of us, our research has drawn some new light on the leading edge of robustness on the directed social network.

1. Introduction

With rapid increasing online social network, the network structure of online social networks has become more complicated than before. Analysis and explaining the dynamics and properties of social networks has become an interesting researching task with plenty of applications in social sciences and many other web application scripts. In some social networks, it is very common for some users who decide to leave the network or begin to stop being active in the activities of their community [1]. This phenomenon is also called as quitting or churn and has absorbed much research attention in social networks. And how to analyze and evaluate network robustness and resilience [2–4] after node departure or removal [5] has always been the hot research points [2, 6] in the last decade. And online social network has been classified as a scale-free network for demonstrating the power-law [7] distribution of degree by many famous complex network scientists [4, 8].

Some recent research results have also told us that network nodes which have a large betweenness [8] value are closely related to swift information and material dissemination in a graph [9, 10] which is useful for quick network robustness and node vulnerability assessment. Because the traditional network robustness and node vulnerability assessment theories are based on undirected and nonweighted networks, it is necessary to research on the relationship of resilience of directed social network after node departure

and the latest found rules of complex network which can be found in Figure 1 for Toy example. And it would be helpful to find some new discipline and cast new light on robustness and vulnerability assessment of directed social network.

In detail, our contributions are as follows:

(1) Basing on classic probability graph theory and clustering coefficient definition, we have proposed two new node centralities named *ITG* (information transfer gain) and $IM_p(v)$ (directed node importance). It can be used to measure the robustness and vulnerability in directed networks especially directed social networks which have been seldom tested before.

(2) We have proposed enough experiment results on undirected artificial networks and directed online social networks to make robustness assessment comparison of *ITG* and $IM_p(v)$, which was mentioned scarcely in former-related works.

(3) Our *ITG*- and $IM_p(v)$-based centrality has been proved to be more accurate, efficient, and faster than classical centrality methods such as degree and betweenness with sufficient experiments results for node robustness assessment in directed social networks.

(4) We have firstly proposed a new rigorous proving process of directed node importance centrality $IM_p(v)$ and its implementation method. We have found that, in the more densely connected directed social network, the $IM_p(v)$ node removal strategy is the most harmful to the network connecting structure. And to the best knowledge of us, we attain the varying trend of iteration times k to the marginal difference ε on directed social networks for the first time based on strict mathematical proof.

The outline of the paper is as follows: Section 2 introduces the related work on robustness and resilience of scale-free network which includes some traditional complex network and social network. Section 3 presents the definitions of network structure quantities which we used to evaluate the robustness and vulnerability of network datasets. And our novel *ITG*- and $IM_p(v)$-based directed node centrality will be introduced. Section 4 proposes the experiments results in synthetic signed network and real-life large networks derived from directed online social media and directed human mobile phone calling network with former undirected node centrality measures and our directed *ITG* centrality measure and $IM_p(v)$. Section 5 gives the final conclusion of this article and draws some new light on the future work.

2. Related Work on Network Robustness

There have been many important research studies on vulnerability assessment and robustness assessment in network structures [11, 12] after node departure or network attack and other significant research areas.

The robustness of ER random network and BA scale-free network was analyzed firstly by Albert [13] in 2000. He used the multiple correlation relationship data of l^{-1} (inverse geodesic length), S (size of the largest connected subgraph), and removed node ration under node attack to make vulnerability and robustness assessment. He discovered that collapse of scale-free network may reach a high price because of the selection and removal of a few nodes which play an important role in maintaining the connectivity of network.

Also, in 2000, Cohen and Callaway [10, 14] observed that real networks demonstrating power-law degree distribution are robust against random node removal but easy to crash in case of attacks to high-degree nodes.

Holme et al. [8] used four different removal strategies which will be introduced in Section 4 in 2002, and he found details of the response of networks according to these attacks on vertices and edges. Holme and his research team observed that the removals by the recalculated degrees and betweenness centralities are always more harmful than the attack strategies based on degrees and betweenness centralities of the initial network. But they only use one real communication network and did not propose some application points.

Then, some researchers have discovered a few valuable results on directed networks. Xu and Wang [6] have done some experiments on the cascading crash on weighted complex networks in 2008. Newman and Ghoshal [2] found that removal of some special single node in the network may cause the bicomponents in the graph to be disconnected.

Malliaros and Vazirgiannis [4] proposed a model to capture this cascading effect node departure of social network, based on engagement dynamics of social networks in 2015. Fragkiskos introduced a new concept of robustness assessment method under cascades triggered by the quitting of nodes based on their engagement level. His results indicated that social networks are very robust and strong under cascades triggered by randomly selected nodes but highly vulnerable in cascades caused by targeted departures of nodes with high engagement level.

In 2016, Gao et al. [10] developed a set of analytical tools with which to identify the natural control and state parameters of a multidimensional complex system for stimulation, and it can attain effective one-dimensional dynamics that can accurately predict the system's resilience. Their proposed analytical framework tool can systematically separate the roles of the system's dynamics and topology, collapsing the behavior of different networks onto a single universal resilience function.

But in the related research mentioned above, these researchers did not model on the information transfer process of directed social networks and cannot fulfill the network robustness and vulnerability assessment requirement of directed social networks nowadays, and it needs to construct some new model and methods to bridge the requirement gap in some efficient ways.

3. Definitions of Network Structure Quantities and Node Centrality Measures

In a scale-free network which includes undirected and directed ones, after some very important nodes are intentionally selected and removed, the network would suffer a serious collapse. It is necessary to research on the network

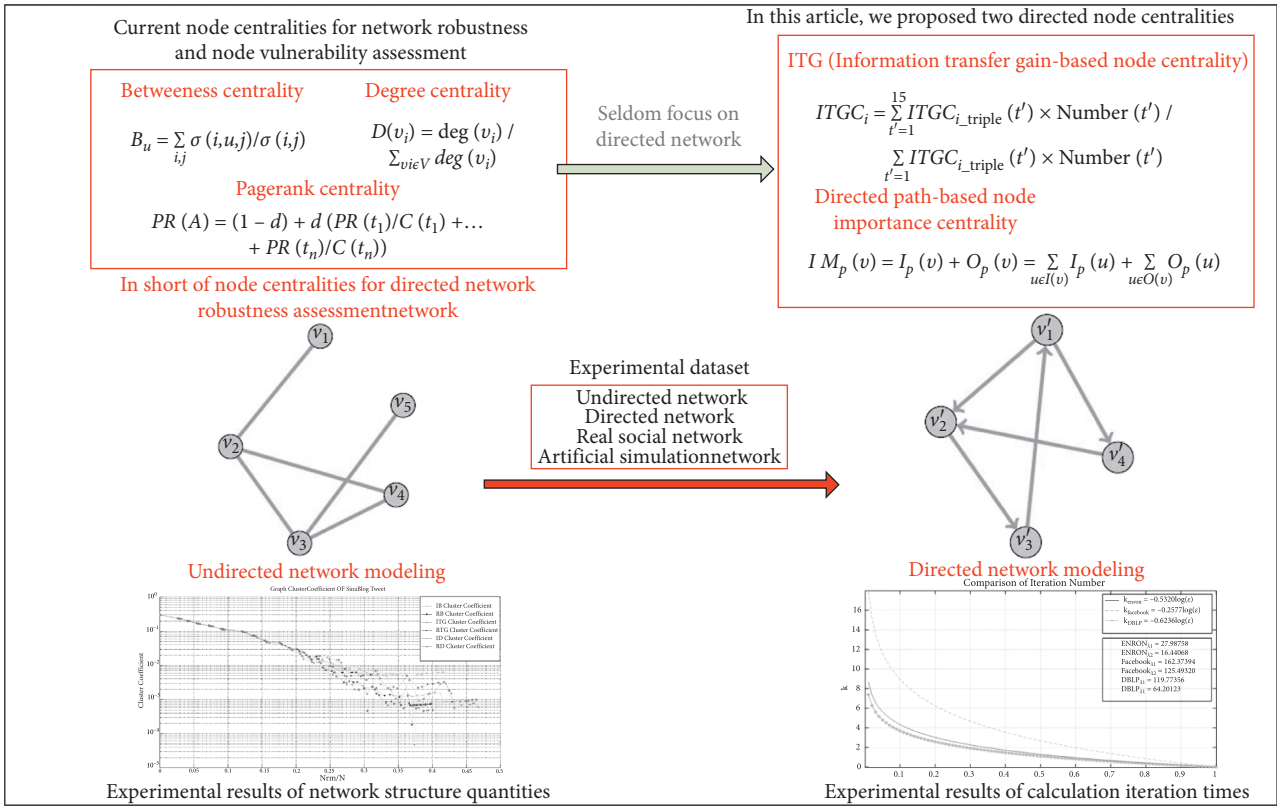


Figure 1: Toy example figure.

resilience in directed networks, so we adopted three key performance quantitative indicators and some famous centralities indicators to make robust assessment of simulated and real networks.

3.1. Network Structure Quantitative Indicator Definition. In this paper, we used two types of network which are undirected and directed networks. On the one hand, for the undirected networks, the network model is unweighted and undirected, which can be demonstrated as $G = (V, E)$. V is the set of nodes with number $N = |V|$, and E is the set of edges with number $L = |E|$. On the other hand, the directed network model is always defined as $\overrightarrow{G} = (V, E)$ with each edge having its own weight and direction to supply the directed vivid information and material spreading.

3.1.1. Average Inverse Geodesic Length. In an undirected network, average inverse geodesic length l is the important network structure quantities after node failure. And in the directed network especially in social networks, the average inverse geodesic length l can be calculated by removal of the direction of edges:

$$l \equiv \langle d(v,w) \rangle \equiv \sum_{v \in V} \sum_{w \neq v \in V} d(v,w). \quad (1)$$

In formula (1), $d(v,w)$ stands for the geodesic path length between different nodes v and w. In traditional social networks of human relationship which is a small-world network, l is always around 6. In the new social networks such as Twitter and Facebook, it would decline to 4 which has been proved by Robert and Sebastiano in Laboratory for web algorithm [15] in 2015. After some nodes are removed, if there is no path between nodes v and w, $d(v,w)$ would reach $+\infty$. And there is another length quantity l^{-1} instead of l:

$$l^{-1} \equiv \langle d(v,w) \rangle \equiv \frac{1}{N(N-1)} \sum_{v \in V} \sum_{w \neq v \in V} \frac{1}{d(v,w)}. \quad (2)$$

The value of $1/d(v,w)$ is zero where there does not exist path from node v to node w.

3.1.2. Network Average Cluster Coefficient. In most network models, the node cluster coefficient C_i reflects the density of connection around some focus nodes. In the whole network, the network average cluster coefficient $C_{G=(V,E)}$ demonstrates the macroscopically cluster characteristics of the whole network to reveal the density of links among neighbor nodes and the cluster coefficient of node can be found in the following formula:

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (3)$$

where k_i stands for stands for the neighbor node number of node i and E_i stands for the actual number of existing links among neighbor nodes of node i. Network average cluster coefficient can be found in the following formula:

$$C_{G=(V,E)} \equiv \frac{1}{N} \sum_{v_i \in V, i=1}^{N} \frac{2E_i}{k_i(k_i - 1)}. \quad (4)$$

3.1.3. Relative Size of Largest Connected Subgraph. In some disconnected large networks, there would exist some separated subgraphs and cannot connect to each other. Size of the largest connected subgraph is also called the node number of the largest connected subgraph in the whole graph which is important to reveal inner connectivity characteristics of graph. In the real network, it can assume that if the connected subgraphs set are $g_1, g_2, g_3, g_4, \ldots\ldots, g_n$ in $G = (V, E)$, then definition of the relative size of the largest connected subgraph can be found in the following formula:

$$S = \frac{\text{Max}\left\{\left|V_{g1}\right|, \left|V_{g2}\right|, \left|V_{g3}\right|, \ldots\ldots, \left|V_{gn}\right|\right\}}{|V|}. \quad (5)$$

3.2. Node Centrality Measures. In network, there are many node centrality measures to calculate the importance of node and each node has its own status and influence to its neighbors. Particularly, in the social relations network, it can demonstrate the relations among nodes. Based on node centrality measures, they can be used to analyze the closeness in the each other dependent relations between different nodes. Furthermore, node centrality measures can be used to calculate the role and graph position of every node for information dissemination influence analysis and other advancing applications [16–21]. Among these researches, node centrality is always the focus of analysis and research results of social network researchers [22] to find the role and status of node. Among these well-known node centralities, we have chosen three most common node centrality statistics and proposed two new node centrality measures which can be found as follows.

3.2.1. Betweenness Centrality. Betweenness is a very famous and vital node centrality measure in network statistics computing. Its detailed definition is as follows: in graph G, if there exists a path between every node v_i and v_j while $\forall v_i, v_j \in V, G$ is a connected graph and there must be a shortest graph to connect v_i and v_j which is always called the geodesic path. Thus, we can use the assuming shortest paths number between two vertices to quantify the importance of a vertex or an edge in terms of its betweenness centrality [23] which can be found in the following formula:

$$B_u = \sum_{i,j} \frac{\sigma(i, u, j)}{\sigma(i, j)}. \quad (6)$$

In formula (6), $\sigma(i, u, j)$ is the number of shortest paths between vertices v_i and v_j that pass through vertex u and $\sigma(i, j)$ is the total number of shortest paths between nodes v_i and v_j.

3.2.2. PageRank Centrality. PageRank is a typical link analysis centrality measure invented by Larry Page in 1997 with the purpose of "calculating" its relative importance within the linking data set. And PageRank also can be applied to calculate node centrality. The simple definition of PageRank can be found in formula (7). In our calculation, damping factor d equals 0.85 as a constant value. The calculation formula of PageRank is as follows.

When page number increases and the damping factor is used to recalculate the PageRank mark of every page, damping factor stands for the real mark when one page is linked to the other page with the range from zero to 1. In our calculation, the damping factor value equals 0.85 as a constant value. The calculation formula of PageRank is as follows:

$$PR(A) = (1 - d) + d\left(\frac{PR(t_1)}{C(t_1)} + \cdots + \frac{PR(t_n)}{C(t_n)}\right). \quad (7)$$

In formula (7), $PR(t_1)$ stands for PR (PageRank) marks in which t_1 brings to page A (t_1 is linked to A). In the same way, $C(t_1)$ stands for pages which are linked to page t_1.

3.2.3. Degree Centrality. In social network, the degree centrality is that central users have the most ties to other actors in the network. If the central user is more powerful and influential in the connecting network, it will have much more links to other users including in and out links. The number of adjacent edges a node has is defined as a degree. In graph G, if the degree of node v_i is $\deg(v_i)$ and the total degree of nodes in the graph is $\sum_{v_i \in V} \deg(v_i)$, the degree centrality of node v_i is $D(v_i)$ which defined in as follows [24]:

$$D(v_i) = \frac{\deg(v_i)}{\sum_{v_i \in V} \deg(v_i)}. \quad (8)$$

From the definition above in formula (8), it can be found that node degree centrality expressed the ratio of degree of the node u to the total degree in the whole graph and some nodes with a higher degree centrality reflect that this node may have some powerful role linking to other node in the network and it may be the most vital person living in the focus of attention.

3.2.4. Directed Path-Based Node Importance Centrality. With the rapid development of information technology, our life is becoming more and more networked. The graph information is closely related to our daily life, such as the directed WeChat relation network and directed Facebook relation network with such complex structures. And most important of all is that the scale of these graphs is enormous and it is hard to handle these so complex and massive graphs. In a directed graph network, how to quantitatively analyze and calculate the importance of each vertex has become an important problem to be solved urgently for those directed graph networks, while they all have some common features [25, 26]. Among these statistical indicators for directed graph networks, directed path-based directed

node importance centrality has become more prominent for its easy and fast calculation.

In the directed graph network G, if there exists a path $e_i = (v_i, v_j)$ between nodes v_i and v_j, directed path-based directed node importance centrality of node v_j can be expressed by the paths across node v_j. The more the across paths are, the more important the node v_j is. Directed path-based node importance centrality $IM_p(v)$ of node v has been posed in the famous book with the name "Computer Science Theory for the Information Age" by John Hopcroft and Kannan [27], but this book did not demonstrate some fast and efficient computing methods of $IM_p(v)$ which we have found in formula (9). And definition of directed path-based node importance centrality $IM_p(v)$ of node v can be found in formula as follows:

$$IM_p(v) = I_p(v) + O_p(v) = \sum_{u \in I(v)} I_p(u) + \sum_{u \in O(v)} O_p(u). \tag{9}$$

In formula (9), $I_p(v)$ stands for the directed paths number which finally ends in node v where $I(v) = \{v | (u, v) \in E\}$ and $O_p(v)$ stands for the directed paths number which starts from node v where $O(v) = v | (u, v) \in E$. Basing on formula (9), we can calculate the $IM_p(v)$ of nodes in directed networks when we solve two critical problems first. One problem is that, in complex structured graphs with huge number nodes, the time complexity of the iterative calculation will be very high and the other problem is that the convergence of the calculation process of formula (9) cannot be reached if we do not use some approximate calculation methods.

For $I_p(v) = \sum_{u \in I(v)} I_p(u)$ in formula (9), it can be converted to iterative computation in the following formula:

$$I_p(v)_k = \sum_{u \in I(v)} I_p(u)_{k-1}. \tag{10}$$

In formula (10), $I_p(v)_k$ stands for the k_{th} iteration calculation result, and for any node v, the initial value of $I_p(v)_k$ is $I_p(v)_0 = 1$. For the reason that computation of formula (10) cannot converge by its diverging computation process, the iterative results of formula (10) need to be standardized in formula (11a) in which α is the damping factor with the value range of $0 < \alpha < 1$ just like the damping factor in the PageRank computation model:

$$I_p(v)_k = (1 - \alpha) + \alpha \cdot \frac{\sum_{u \in I(v)} I_p(u)_{k-1}}{\sqrt{\sum_{u \in I(v)} \left(I_p(u)_{k-1}\right)^2}}. \tag{11a}$$

For the similar computation principle, the standardized form of $O_p(v)$ in formula (9) can be written as follows:

$$O_p(v)_k = (1 - \alpha) + \alpha \cdot \frac{\sum_{u \in I(v)} O_p(u)_{k-1}}{\sqrt{\sum_{u \in I(v)} \left(O_p(u)_{k-1}\right)^2}}. \tag{11b}$$

In computation of formulas (11a) and (11b), we can use M to stand for the adjacency matrix of directed graph network G and I_p to stand for the vector composed by all the $I_p(v)$ values of the node set $v \in V$ in G while $|v| = n$. And I_p can be expressed by $I_p = [I_p(v_1), I_p(v_2), \ldots, I_p(v_n)]^{\mathrm{T}}$, while formula (11a) can be changed as follows:

$$I_p = (1 - \alpha) \cdot I + \alpha \cdot \frac{M \cdot I_p}{\left\| M \cdot I_p \right\|}. \tag{12a}$$

In formula (12a), $I = [1, 1, \ldots, 1]^{\mathrm{T}}$ and I_p can be initialized to $I_p = I$. For the similar computation principle, formula (11b) can be changed as follows:

$$O_p = (1 - \alpha) \cdot I + \alpha \cdot \frac{M \cdot O_p}{\left\| M \cdot O_p \right\|}. \tag{12b}$$

By using formulas (12a) and (12b), the computation of directed path-based node importance centrality $IM_p(v)$ of node v can be transferred to the computation of adjacency matrix M of directed graph network G with great promotion. Furthermore, we will demonstrate the whole proof process which using the Power Iteration Theory [28] to prove that formulas (12a) and (12b) can converge to a definite vector by finite times of computations. For the adjacency matrix $M = (m_{ij}) \in R^{n*n}$ of directed graph network G, it has the eigenvalue set of $\lambda_1, \ldots, \lambda_n (\lambda_i \in R)$, while $\lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_n|$ and $\lambda_1 = 1$. At the same time, $Mx_i = \lambda_i x_i$ and eigenvectors set $\longrightarrow_{x_i} (i = 1, \ldots, n)$ can satisfy that $\longrightarrow_{x_i} \cdot \longrightarrow_{x_j} = 0 \ (i \neq j)$.

We can assure that $y^{(0)} = \sum_{i=0}^{n} c_i x_i$ and $y^{(k)}$ can be expressed as $y^{(k)} = \sum_{i=0}^{n} c_i \lambda_i^k x_i$.

Here, $[y^{(0)}, y^{(1)}, \ldots, y^{(k)}]$ can be used to replace $I_p = [I_p(v_1), I_p(v_2), \ldots, I_p(v_n)]^{\mathrm{T}}$ or $O_p = [O_p(v_1), O_p(v_2), \ldots, O_p(v_n)]$ for formulas (11a) and (11b).

For $y^{(i)} \in \{y^{(0)}, y^{(1)}, \ldots, y^{(k)}\}$, $0 \leq i < k$,

$$\begin{aligned} \left\| y^{(k+1)} - y^{(k)} \right\| &= \left\| \sum_{i=1}^{n} c_i \lambda_i^{k+1} x_i - \sum_{i=1}^{n} c_i \lambda_i^k x_i \right\| = \left\| \sum_{i=2}^{n} c_i \lambda_i^k (\lambda_i - 1) x_i \right\| \\ &\leq \sum_{i=2}^{n} \left\| c_i \lambda_i^k (\lambda_i - 1) x_i \right\| \leq (n-1) |c|_{\max} |\lambda_2 - 1| |\lambda_2|^k \|x_i\|_{\max} \leq a |\lambda_2|^k, \end{aligned} \tag{13a}$$

where $a > 0, a = (n-1)|c|_{\max}|\lambda_2 - 1|\|x_i\|_{\max}$.

We can also prove that the iteration times k ($k = [1, N]$ of computation for $\|y^{k+1} - y^k\|$) are around a constant level value in the following formula:

$$\forall \left\| y^{(k+1)} - y^{(k)} \right\| < \varepsilon \Longrightarrow a|\lambda_2|^k < \varepsilon (0 < \varepsilon < 1) \Longrightarrow k \log|\lambda_2| < \log \varepsilon - \log a \Longrightarrow k > \frac{\log \varepsilon - \log a}{\log|\lambda_2|}. \tag{13b}$$

And the constant level value of k is

$$k > \eta \log \varepsilon + \beta \left(\eta < 0, \eta = \frac{1}{\log|\lambda_2|}, \beta = -\frac{\log a}{\log|\lambda_2|} \right). \tag{13c}$$

3.2.5. ITG-Based Directed Node Centrality. In most actual social networks such as Twitter, Facebook, and WeChat, the two friends of someone can be friends with each other. And this phenomenon can be found in both undirected networks and directed networks. So this attribute of social networks sometimes is called the clustering characteristic of network [7]. Network average clustering coefficient reflects the microscopically clustering characteristic of network and has become the very important measure of adjacent nodes which are connected closely. The definition of network node clustering coefficient can be found in Section 3.1.2. Based on the classic probabilistic graphical model (PGM) theory from Turing Award Owner Pearl [29], we have made a detailed research on directed node influence clustering coefficient [30] to propose a new vector influence clustering coefficient model with both information propagation direction and information propagation probability. The new vector influence clustering coefficient model starts from the two basic directed triple forms of vertex i in the directed graph which can be found in Figure 2.

Basing on deducting from Figures 2(a) and 2(b), we can get the related edge information propagation probability and direction of different two types including 36 subgraphs of all different triples [30]. And then we can calculate all numerical values of directed *ITG* node centrality in directed social networks in all kinds of situations by the following formula:

$$\sum_{i \longleftrightarrow j} ITG = \underset{i \longleftrightarrow j}{ITG} + \underset{i \longleftrightarrow k \longleftrightarrow j}{ITG} = \underset{i \longleftrightarrow j}{ITG} + \underset{i \longleftrightarrow k}{ITG} + \underset{k \longleftrightarrow j}{ITG}, \tag{14}$$

$$\sum_{i \longrightarrow k} ITG = \underset{i \longrightarrow k}{ITG} + \underset{i \longleftrightarrow j}{ITG} + \underset{j \longrightarrow k}{ITG}. \tag{15}$$

Because there are three possible directional statuses for each edge in Figures 2(a) and 2(b), the adjacent edge of node i has three different definitions, which are friends relationship ($i \longleftrightarrow j$), following relationship ($i \longrightarrow j$), and fan relationship ($i \longleftarrow j$). At the same time, the opposite edge $i \longleftrightarrow k$ of node i also has three definitions. Furthermore, the edge $j \longleftrightarrow k$ has three types of relationships, in which node j and node k are friends, node j follows node k, and node k follows node j. We can use 0, 1, and 2 to stand for the relationships and substitute the three different definitions, and we obtain the following 27 arrangements in Table 1 [30].

In the above 27 arrangement cases, because node i is the source node, we can find some symmetry results, and finally, we can get 15 independent results. And the *ITG*-based directed node centrality value from information transfer gain clustering coefficient (*ITGC*) of node i in a directed network can be finally summed by the 15 different independent results in formula (16) as follows:

$$ITGC_i = \frac{\sum_{t=1}^{15} ITG_{i\_\text{triangle}}(t) \times \text{Number}(t)}{\sum_{t'=1}^{6} ITG_{i\_\text{triple}}(t') \times \text{Number}(t')}. \tag{16}$$

$ITGC_i$ is the *ITG* value of node i in a directed network. $\sum_{t=1}^{15} ITG_{i\_\text{triangle}}(t) \times \text{Number}(t)$ is the weighted number of triangles which use node i as the top vertex (i.e., the information transfer source node), and its weight is the *ITG* (information transfer gain) contribution $ITG_{i\_\text{triangle}}(t)$ from the 15 different types of weighted triangles multiplied by its counted number $\text{Number}(t)$. $\sum_{t'=1}^{6} ITG_{i\_\text{triple}}(t') \times \text{Number}(t')$ is the weighted number of the triples using node i as the top vertex; its weight is the weighted sum of the six $ITG_{i\_\text{triple}}(t')$ values of different types of triples multiplied by its counted number $\text{Number}(t')$. Similar to undirected clustering coefficients, *ITGC* has the same characteristic, measuring the tightness of the graph to form tight communities.

4. Experiments on Network Robustness Assessment

4.1. Simulation of Node Failure Generation Method. Basing on the importance and connectivity of different nodes in the network, there are eight node attack strategies [8] always chosen by researchers to evaluate network robustness such as follows:

(1) The ID removal strategy: the attack starting from the node with the highest degree and node attack strategy uses the initial node degree distribution.

(2) The IB removal strategy: the attack starting from the node with the highest betweenness and node attack strategy uses the initial node betweenness distribution.

(3) The *ITG* removal strategy: the attack starting from the node with the highest *ITG* centrality and node attack strategy uses the initial node *ITG* centrality distribution.

(4) The *IMP* removal strategy: the attack starting from the node with the highest $IM_p(v)$ centrality and

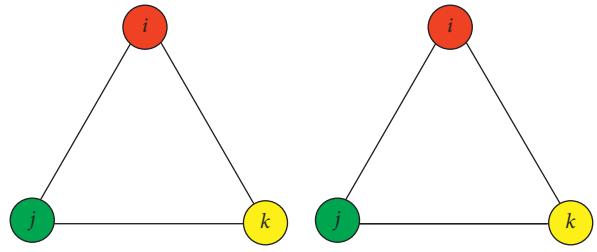


Figure 2: Two triple forms of vertex i in undirected graph.

node attack strategy uses the initial node $IM_p(v)$ centrality distribution.

(5) The RD removal strategy: using the recalculated node degree distribution at every removal step.

(6) The RB removal strategy: using the recalculated node betweenness at every step.

(7) The RTG removal strategy: using the recalculated node *ITG* centrality distribution at every removal step.

(8) The RIMP removal strategy: using the recalculated node $IM_p(v)$ centrality distribution at every removal step.

4.2. Experiment Result. Selection of typical and persuasive dataset is very important to experiment results, and we selected some classical undirected network and directed network dataset including synthetic signed network and real-life large networks used by Albert et al. [13] and Holme et al. [8]. And these typical datasets can be used to validate the network robustness and node vulnerability assessment in undirected and directed networks.

From Figures 3 to 21, all the prominence in experiment result which was created by *ITG*, *RTG*, *IMP*, and *RIMP* strategies have be pointed out by red ellipses and text tag. Based on these marks, we can easily find the more strong effects by our proposed *ITG* and $IM_p(v)$ centralities.

4.2.1. Undirected Dataset and Experiment Results. In Table 2, the network dataset includes the classical BA scale-free network (a) proposed by Albert et al. [13] (generation parameter is $m_0 = 5$, $m = 4$, $p_t = 0.8$, and $n = 490$) and the undirected call community [31] graph (b) from cellphone calling records in one month in China of a southern city, the LFR (Lancichinetti Fortunato Radicchi) [32] benchmark network with generation parameter of $N = 1000$, $k_{\text{degree}} = 2$, $C_{\min} = 20$, $C_{\max} = 100$, $u = 0.3$, and $C_{\text{degree}} = 1$. For the reason that the LFR benchmark has presented a much solid testing dataset for algorithms and having good performance from other dataset, we used LFR benchmark to generate testing dataset having typical attributes compared with real networks, such as real node degree distribution and heterogeneous distribution of community size. Because the above networks are important undirected networks, the three network datasets are chosen for our research.

Table 1: Permutation and combination of 27 conditions.

000	001	002	010	011	012	020	021	022
100	101	102	110	111	112	120	121	122
200	201	202	210	211	212	220	221	222

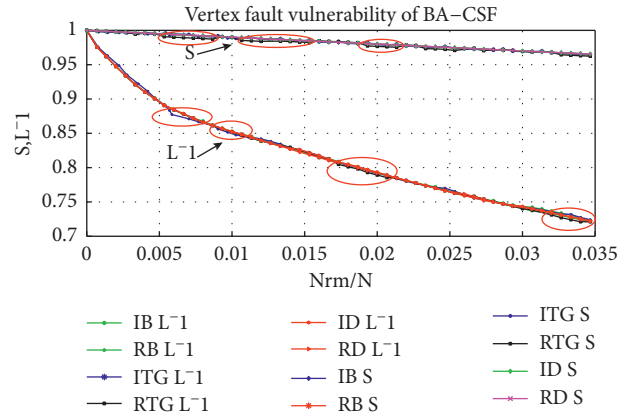


Figure 3: l^{-1}&S result of BACSF.

Now we will give the brief introduction of vertical and horizontal coordinates of the graph, in which the x-axis in Figures 3 to 8 stands for the removed node ration (Nrm: the number of removed nodes) to initial graph. The y-axis stands for the relative value of S and l^{-1} (the ration of S and l^{-1} after every step of node removal to initial S and l^{-1} in the network). Figures 7 and 11 consist of 400 calculated results each in 200 networks generated by the four simulated node generation mechanisms above to avoid random deviation (every mechanism generated 50 fault nodes).

In calculation of *ITG*-based node coefficient in undirected networks, all the information transfer gain probability a between nodes i and j for undirected edges will be equal to 0.5 and we would calculate *ITG* node centrality coefficient of each node and the whole graph.

4.2.2. Experiment Results Analysis for Undirected Dataset. After processing and analyzing the experimental data above, it can be clearly found in Figures 3, 5, and 7 that the influence to l^{-1} is as follows:

(1) When the number of removed nodes is very small that is to say when $0.0 < N_{rm}/N < 0.005$, it is the very

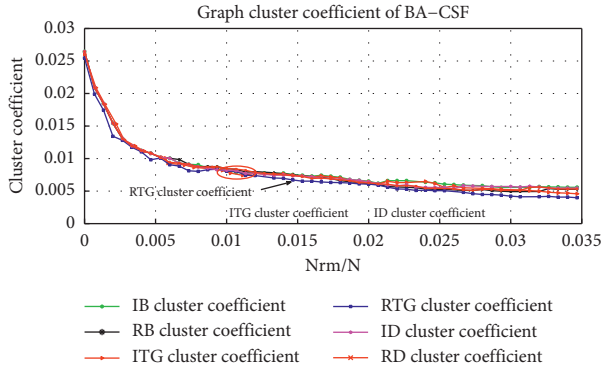


Figure 4: $C_{G=(V,E)}$ result of BACSF.

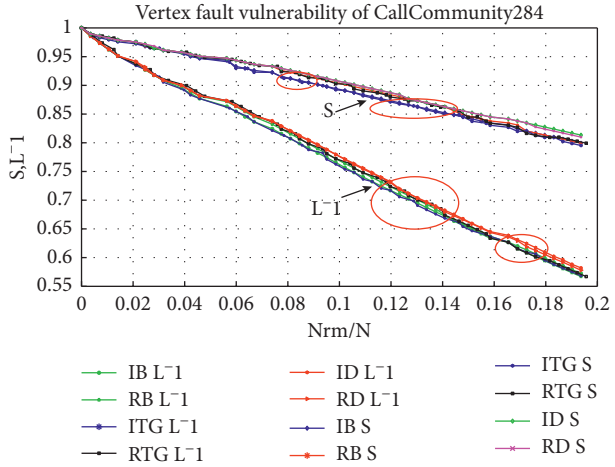


Figure 5: l^{-1}&S result of CallCommunity.

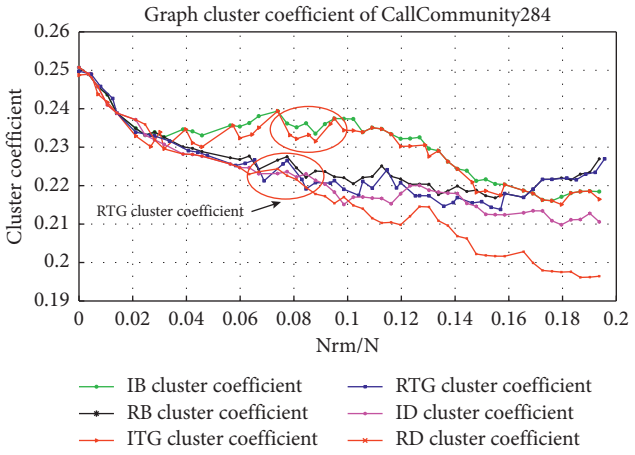


Figure 6: $C_{G=(V,E)}$ result of CallCommunity.

early stage of collapse time by node removal of the six different removal strategies and the l^{-1} curve falls quickly to make the whole graph shrink with a high speed.

(2) During the fall process of l^{-1}, the *RTG* curve, *RB* curve, and *RD* curve look like falling faster than other three removal strategies.

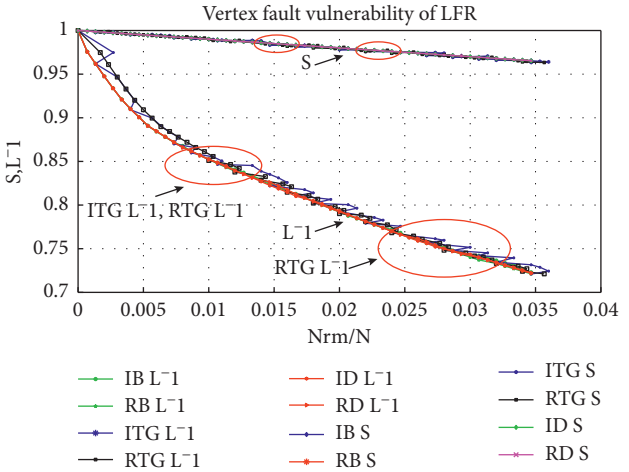


Figure 7: l^{-1}&S result of LFR.

(3) Among the six different curves of S, the curves of *ITG*, *IB*, and *ID* are more likely a linear curve than other ones, and the experimental result has shown the correctness of *ITG* in its almost linear curve with *IB* and *ID*.

In the common circumstances, the network will be more denser while having the larger value. But in generally, the value of network average cluster coefficient having a sudden increase after the continuous decreasing trend would tell that the node removal can make the network suddenly into disconnected subnetworks. Besides that, Albert et al. [13] and other researchers have also found that when the node having a very high betweenness value leaves the network, it may trigger the huge collapse in the whole network on the system level with a sudden unpredicted speed. Figures 4, 6, and 8 demonstrate the corresponding variation trend of $C_{G=(V,E)}$:

(1) It can be also clearly found in Figures 4, 6, and 8 that, among the six curves of S, the *RTG, ITG, RB, and IB* removal strategies are more harmful than *RD* and *ID* where in some datasets, the *RTG* removal strategy is much more harmful than others such as in S and $C_{G=(V,E)}$ curve of LFR in Figures 7 and 8.

(2) There have been much more rises in *ITG* and *IB* curves than other curves for the reason that it can break the whole network into disconnected subnetworks with much higher numbers. It has also proved that our *ITG* node centrality is a very good node centrality by the performance of curve S, $C_{G=(V,E)}$, and l^{-1}.

4.2.3. Directed Dataset and Experiment Results. Most social networks are directed networks with directed edges such as e-mail networks, twitter network, and calling record. We selected some representative social network dataset in Table 3. The OSLOM (a) dataset was provided by the open source algorithm OSLOM [33] as an example dataset of directed social networks. The subject reference (b) dataset

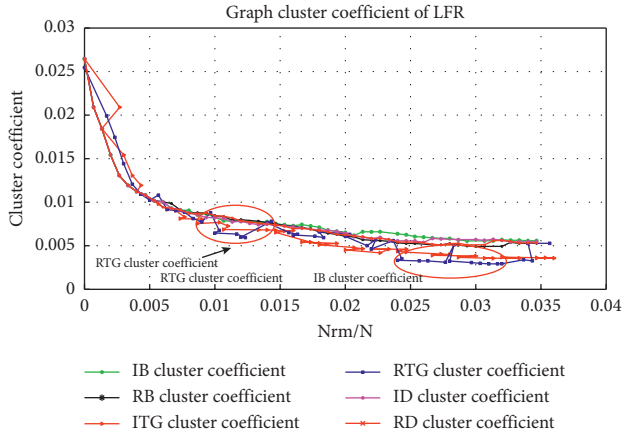


Figure 8: $C_{G=(V,E)}$ result of LFR.

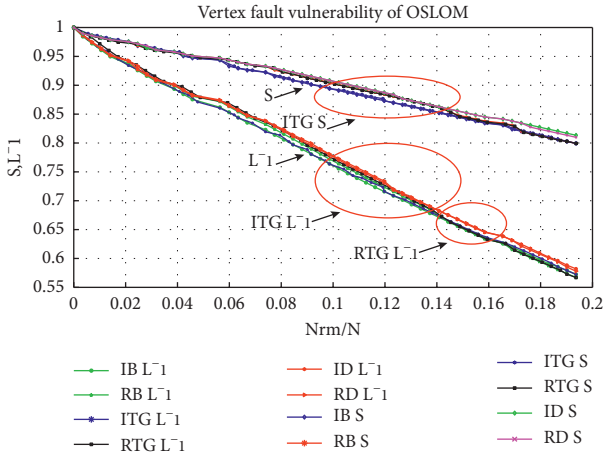


Figure 9: l^{-1}&S result of OSLOM.

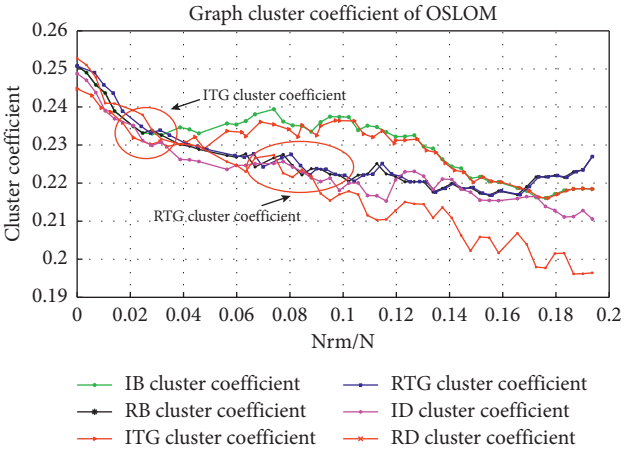


Figure 10: $C_{G=(V,E)}$ result of OSLOM.

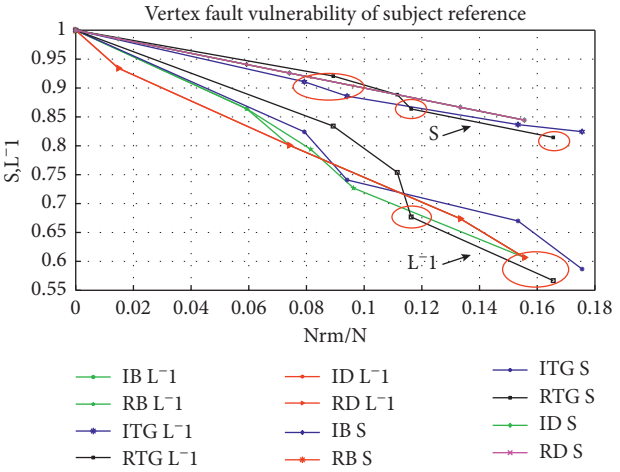


Figure 11: l^{-1}&S result of subject reference.

was provided by the INFOMAP algorithm [34] which was a subject reference network from the research of Physics, Chemistry, Biology, and Ecology in 6,128 journals connected by 6,434,916 citations. The SinaBlog Tweet (c) dataset came from the tweeting message reposting chain of one famous Chinese scholar in SinaBlog. The Calling Record (d) dataset was provided by cellphone calling records from another city in one month in China. All the above datasets are typical directed social networks.

The x-axis in Figures 9 to 16 stands for N_{rm}. The y-axis stands for the relative value of S and l^{-1}. Specially, all the network datasets in Table 3 would be processed as undirected graph in calculation of node betweenness and degree for *IB, RB, ID,* and *RD* strategy besides *ITG* and $IM_p(v)$ strategy. But, all the network datasets in Table 3 would be processed as directed graph in calculation of *ITG* node centrality coefficient. For the reason that there are only 40 nodes in subject reference (b) dataset, we only calculated the top five nodes.

(1) Experiment Results *of ITG with* $IM_p(v)$ *to IB, RB, ID, and RD.*

(2) Experiment Results Analysis for Directed Dataset of ITG with $IM_p(v)$ *to IB, RB, ID, and RD.* Figures 9 to 16 display the experiment results on directed social network, and we can find that the *ITG*, IMP, RTG, and RIMP strategies have amplified their harm in directed networks:

(1) l^{-1} and S: *ITG, IMP, RTG,* and *RIMP* are most harmful strategies, and the *RIMP* strategy is more powerful than *RTG* especially in the late stage of Figures 11, 13, and 15.

(2) $C_{G=(V,E)}$: among the six $C_{G=(V,E)}$ curves, the sudden rises caused by *IMP*, *ITG*, and IB are much more than the rises caused by other strategies where *IMP* is the most, where in Figure 14, we used the Y-axis logarithmic to show details more clear. The influence to $C_{G=(V,E)}$ approximately has shown that *RIMP* > *IMP* > *ITG* > *RTG* > *IB* > *RB* > *ID* > *RD*.

(3) Especially in the sparse directed network of subject reference, *IMDP* L^{-1}, *ITG* L^{-1}, *RTG* l^{-1}, and *RTG CC* curves in Figures 11 and 12 finally reached the finest experiment result and caused the biggest cascade. In Figures 13 and 14 of the SinaBlog Tweet network, the *RIMP* L^{-1}, *RIMP CC*, *RTG* l^{-1}, *RTG S*, and *RTG CC*

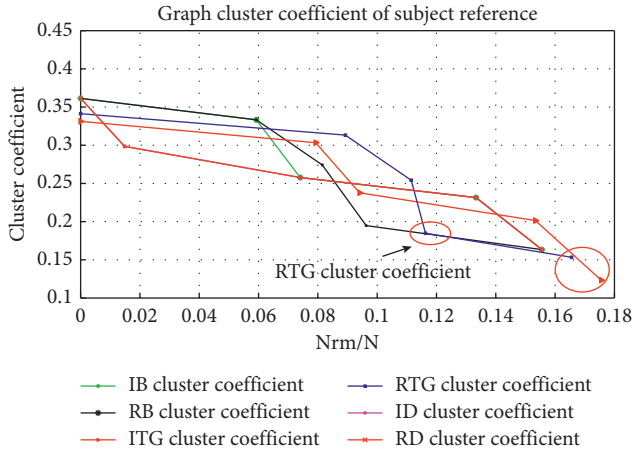


Figure 12: $C_{G=(V,E)}$ result of subject reference.

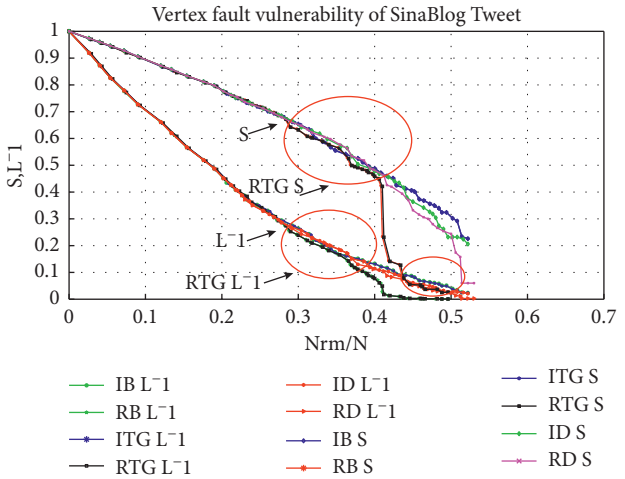


Figure 13: l^{-1}&S result of SinaBlog Tweet.

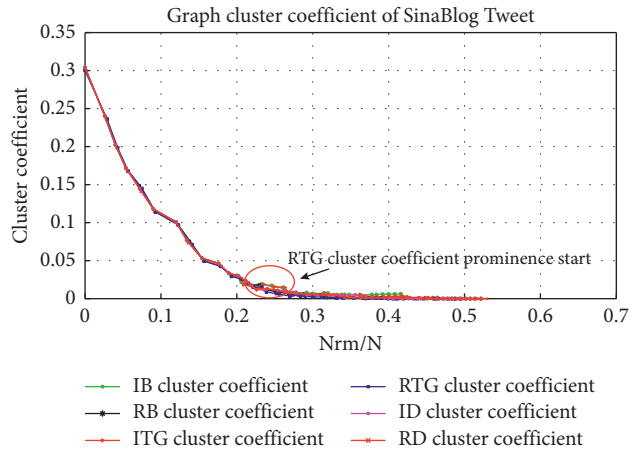


Figure 14: $C_{G=(V,E)}$ result of SinaBlog Tweet.

curves are the best result of experiment to other curves. In dense network of Calling Record, the *RIMP S*, *RIMP CC*, *RTG S*, and *RTG CC* also showed the power of $IM_p(v)$ and *ITG* node centrality by the leading experiment values.

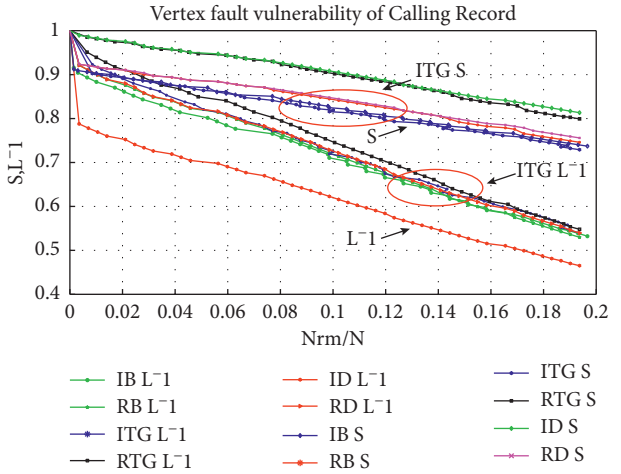


Figure 15: l^{-1}&S result of Calling Record.

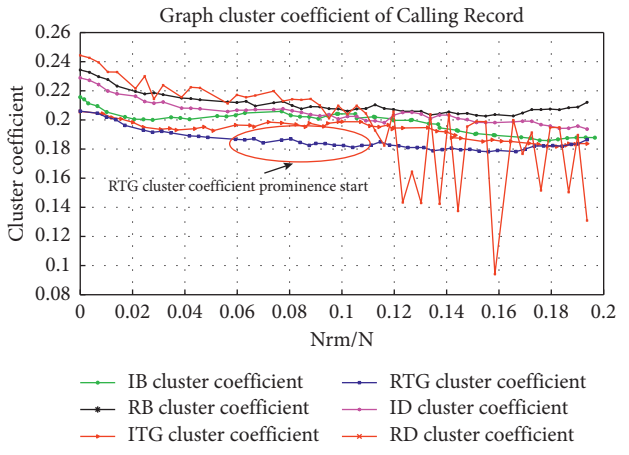


Figure 16: $C_{G=(V,E)}$ result of Calling Record.

Then, we can summarize from Figures 3 to 16 that the $IM_p(v)$- and *ITG*-based strategies have good experiment result in undirected and directed networks, and it has been proved for its correctness.

(3) Directed Experiment Dataset of ITG to $IM_p(v)$. In this part, we added two typical directed networks with much nodes inside which can be found in Table 4. The first one is the DBLP directed network dataset which is a famous directed heterogeneous information network that contains a dataset of author-centric English literature in the field of computer science with 14736 papers and 14475 authors [28]. We selected the data related to the field of computer, including database, data mining, artificial intelligence, and information retrieval including titles of papers published in various fields, authors who published more than five papers, abstracts of papers, and conferences titles. The another directed network dataset is the well-known ENRON e-mail network among employees of ENRON company from May 11, 1999, to May 21, 2002 (http://www.cs.cmu.edu/~/enron/). This ENRON e-mail network is divided into small parts by time stamp in each seven days, and the whole network is composed by 150 persons and 1526 emails among them [35, 36]. Its network

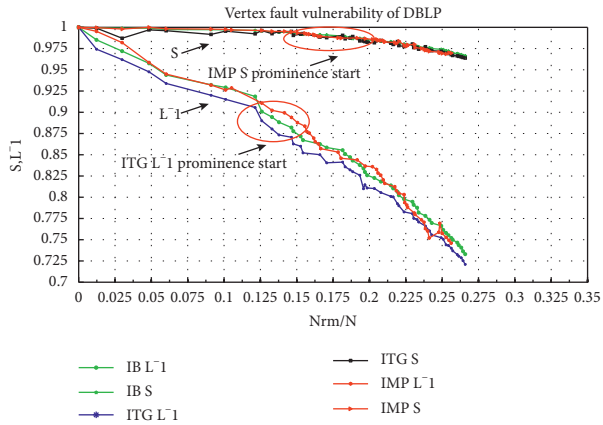


Figure 17: l^{-1}&S result of DBLP.

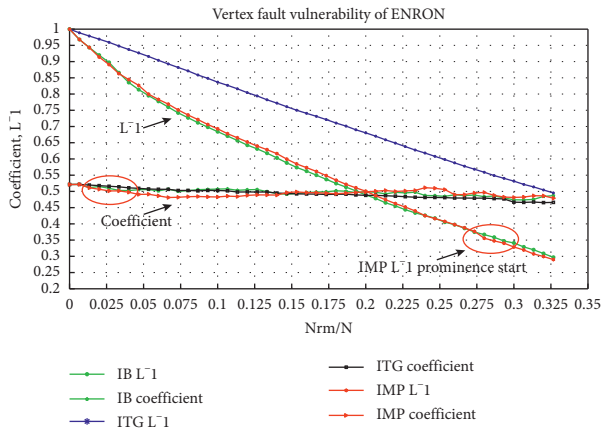


Figure 18: $l^{-1}\$C_{G=(V,E)}$ result of ENRON.

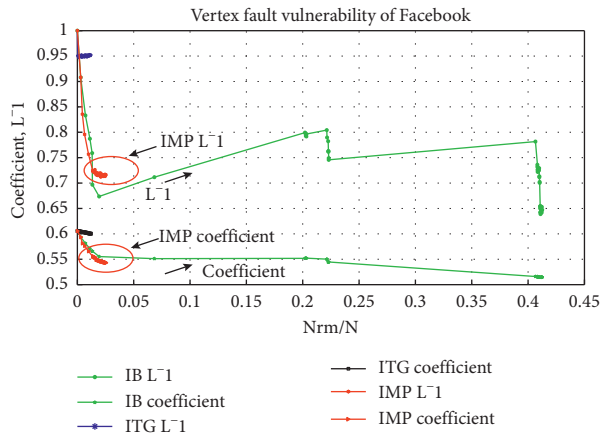


Figure 19: l^{-1}&$C_{G=(V,E)}$ result of Facebook.

structure clearly shows some important moments in the company's development, such as the company's collapse and the suicide of its former CEO. Dataset C is a Facebook dataset which was collected from survey participants using Facebook app. The dataset includes node features (profiles), circles, and ego networks [37]. The Facebook dataset has been

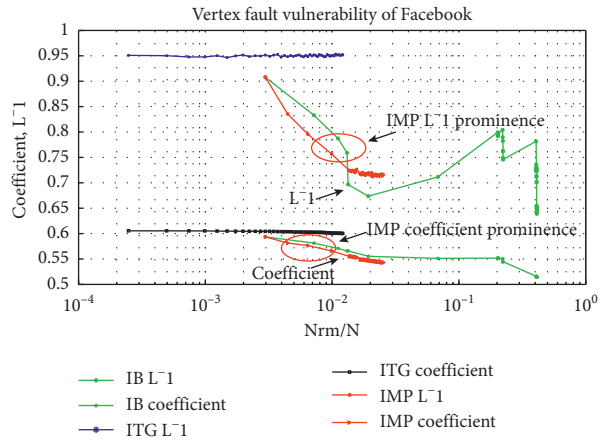


Figure 20: l^{-1}&$C_{G=(V,E)}$ result of Facebook (X-axis logarithmic coordinate).

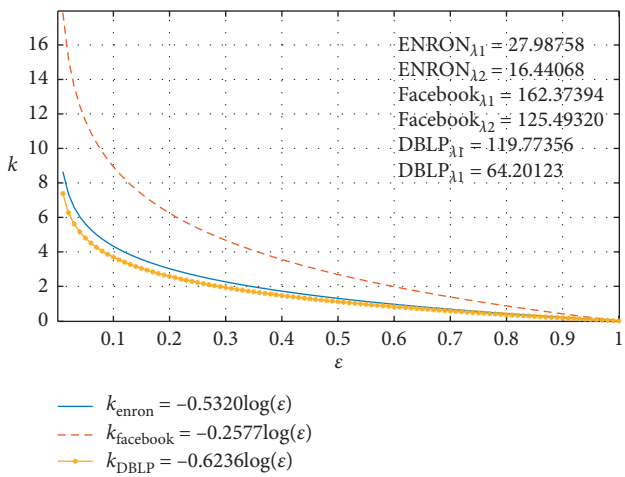


Figure 21: l^{-1} comparison result of iteration times.

anonymously processed by replacing the Facebook-internal ids for each user with a new value.

(4) Experiment Results for Directed Dataset of ITG to $IM_p(v)$.

Figures 17 and 18 display the experiment results of l^{-1} and coefficient on directed social network DBLP and ENRON. For the reason of the small scale of ENRON network dataset and very little change in when deleting the top 50 nodes, we only offered the l^{-1} and coefficient result of it.

It is obviously that, in these directed networks, the influence of IB, ITG, and $IM_p(v)$ has demonstrated a different varying trend in the undirected networks from Table 1. The computation of average inverse geodesic length l always cause high computing costs, while it costs us about 2.4 billion times shortest path search in the dataset of Facebook which has 4039 nodes and 88234 edges ($4000 * 4000 * 3 * 50 = 2.4 * 10^9$) and its computation result only can make up the three curves of $IB\,l^{-1}$, $ITG\,l^{-1}$, and $IMP\,l^{-1}$.

Table 2: Undirected network dataset.

	BACSF (a)	Call Community (b)	LFR (c)
Number of node N	490	284	1000
Number of link L	1892	3030	15384
Edge density (L/N)	3.8612	10.6690	15.3840

N: number of nodes; K_{degree}: degree power-law distribution index; $C_{\min}$: minimum number of nodes; $C_{\max}$: maximum number of nodes; u: mix hybrid parameter; C_{degree}: community size power-law distribution index.

Table 3: Directed network dataset.

	OSLOM (a)	SubjectReference (b)	SinaBlogTweet (c)	CallingRecord (d)
Number of node N	301	40	1492	286
Number of link L	6234	306	1490	3934
Edge Density (L/N)	20.7110	7.6500	0.9987	13.7552

Table 4: Another directed network dataset.

	DBLP (a)	ENRON (b)	Facebook (c)
Number of node N	14376	150	4039
Number of link L	14475	1526	88234
Edge Density (L/N)	1.0069	10.1733	21.8455

Figures 19 and 20 demonstrate the experiment results of l^{-1} and coefficient on directed social network Facebook. In order to demonstrate more details, the x-axis in Figure 20 used the logarithmic coordinates.

Figure 21 demonstrates the comparison result of iteration times in different datasets of Table 4 when computing the directed path-based node importance centrality $IM_p(v)$. By calculating formula $\Longrightarrow k > ((\log \varepsilon - \log a)/\log|\lambda_2|)$ (13b), we attain the varying curve of iteration times k to the marginal difference ε, while ε can be calculated by $\varepsilon \geq \|y^{(k+1)} - y^{(k)}\|$. Furthermore, each largest eigenvalue value λ_1 and second largest eigenvalue value λ_2 of dataset DBLP, ENRON, and Facebook also can be computed.

(5) Experiment Results Analysis for Directed Dataset of ITG to $IM_p(v)$. Figures 16 to 21 display the compared experiment results of *IB*, *ITG*, and $IM_p(v)$ node removal strategies on directed social networks. In addition, we explore the varying trend of iteration times k to the marginal difference ε.

(1) Generally after analyzing the demonstrated data in Figures 17 and 18, we can find that the influence of node with high *ITG* values have the most harmful effect to the l^{-1} in directed networks with low edge density such as *DBLP*. But in directed networks with high edge density such as ENRON, *IB* strategy and *ITG* strategy are just like doing the same effect.

(2) But in the directed network Facebook with a more higher edge density, we can find that the *ITG* node removal strategy suddenly lost its magic and the $IM_p(v)$ removal strategy does the best harmful effect to the Facebook network, while the *IB* removal strategy followed. And it maybe needs to be varied in more densely connected directed social networks such as Twitter and WeChat.

(3) In Figure 21, we can clearly find that the relationship of k and ε. The more densely connected directed network Facebook has the larger iteration times number k. And in dataset Facebook, when ε is more close to zero, iteration times rises to around twenty which has shown high efficiency of our node $IM_p(v)$ computing algorithm.

5. Conclusion

In this paper, we have proved new information transfer gain-(*ITG*-) based probability clustering coefficient and directed node importance centrality $IM_p(v)$ for measuring directed graph. Our comparisons in the variation trend of some key performance quantities of network robustness and node vulnerability assessment are useful and helpful. Comparison could help us to capture this cascading effect in directed online social networks. Experiments results showed that node *RIMP* and *RTG* strategies are more harmful than node betweenness-based strategies such as *RB* and *IB* in directed social networks including real Sina Blogging and Calling Record network. With sufficient experiments in synthetic signed networks and real networks derived from directed online social media and directed human mobile phone calling network, it has been proved that our *ITG*- and $IM_p(v)$-based directed social network robustness and node vulnerability assessment method is more accurate, efficient, and faster than several classical traditional centrality methods such as degree and betweenness. Furthermore, we will carry out our *ITG* centrality on more types and more large-scale directed social networks.

In addition, we propose a new proving process of directed node importance centrality $IM_p(v)$ in Section 3.2.4. By rigorous mathematical derivation and approximate calculation, to the best knowledge of us, we attain the varying trend of iteration times k to the marginal difference ε on directed social networks for the first time to our best knowledge.

Data Availability

These data used to support the findings of this study are included within the supplementary information file.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (Grant no. 2017YFC0820603) and Advance Research Project of Shared Information System in 2019 (Grant no. 31511040103).

Supplementary Materials

The supplementary files contain the data which are used to support the findings of this study. (*Supplementary Materials*)

References

[1] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence*, pp. 4981–4987, Yokohama, Japan, July 2020.

[2] M. Newman and G. Ghoshal, "Bicomponents and the robustness of networks to failure," *Physical Review Letters*, vol. 100, no. 13, p. 138701, 2008.

[3] T. Zhou, W. Bai, B. Wang, Z. Liu, and G. Yan, "A brief review of complex networks," *Physics*, vol. 34, no. 1, pp. 31–36, 2005.

[4] F. D. Malliaros and M. Vazirgiannis, "Vulnerability assessment in social networks under cascade-based node departures," *EPL (Europhysics Letters)*, vol. 110, no. 6, p. 68006, 2015.

[5] M. Subramanian, *Network Management: Principles and Practice*, Pearson Education India, Chennai, India, 2010.

[6] J. Xu and X. F. Wang, "Cascading failures in scale-free coupled map lattices," in *Proceedings of the 2005 IEEE International Symposium on Circuits and Systems*, pp. 3395–3398, Kobe, Japan, June 2005.

[7] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E*, vol. 74, no. 3, p. 036104, 2006.

[8] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, no. 5, p. 056109, 2002.

[9] A. A. Nanavati, R. Singh, D. Chakraborty et al., "Analyzing the structure and evolution of massive telecom graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 703–718, 2008.

[10] J. Gao, B. Barzel, and A.-L. Barabási, "Universal resilience patterns in complex networks," *Nature*, vol. 530, no. 7590, pp. 307–312, 2016.

[11] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[12] J. Wu, S. Pan, X. Zhu, and C. Zhihua, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.

[13] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[14] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Resilience of the internet to random breakdowns," *Physical Review Letters*, vol. 85, no. 21, p. 4626, 2000.

[15] R. Meusel, S. Vigna, O. Lehmberg, and C. Bizer, "The graph structure in the web - analyzed on different aggregation levels," *Journal of Web Science*, vol. 1, no. 1, pp. 33–47, 2015.

[16] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function*, Cambridge University Press, Cambridge, UK, 2010.

[17] D. Garcia, P. Mavrodiev, and F. Schweitzer, "Social resilience in online communities: the autopsy of friendster," in *Proceedings of the First ACM Conference on Online Social Networks*, pp. 39–50, Boston, MA, USA, October 2013.

[18] S. Wu, A. Das Sarma, A. Fabrikant, S. Lattanzi, and A. Tomkins, "Arrival and departure dynamics in social networks," in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pp. 233–242, Rome, Italy, February 2013.

[19] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[20] M. Scheffer, S. R. Carpenter, T. M. Lenton et al., "Anticipating critical transitions," *Science*, vol. 338, no. 6105, pp. 344–348, 2012.

[21] T. Nepusz and T. Vicsek, "Controlling edge dynamics in complex networks," *Nature Physics*, vol. 8, no. 7, pp. 568–573, 2012.

[22] J. S. Carrington, J. Peter, and S. Wasserman, *Models and Methods in Social Network Analysis*, Cambridge University Press, Cambridge, UK, 2005.

[23] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, "Characterization of complex networks: a survey of measurements," *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.

[24] C. Kiss and M. Bichler, "Identification of influencers-measuring influence in customer networks," *Decision Support Systems*, vol. 46, no. 1, pp. 233–253, 2008.

[25] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web," Technical Report, Stanford InfoLab, Stanford, CA, USA, 1999.

[26] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[27] J. Hopcroft and R. Kannan, *Computer Science Theory for the Information Age*, Shanghai Jiao Tong University Press, Shanghai, China, 2013.

[28] M. Ji, Y. Sun, M. Danilevsky, J. Han, and J. Gao, "Graph regularized transductive classification on heterogeneous information networks," in *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 570–586, Springer, Barcelona, Spain, September 2010.

[29] J. Pearl, *Probabilistic Reasoning in Intelligent Systems, Representation and Reasoning*, Morgan Kaufmann, San Mateo, CA, USA, 1988.

[30] X. Deng, J. Zhai, T. Lv, and L. Yin, "Efficient vector influence clustering coefficient based directed community detection method," *IEEE Access*, vol. 5, pp. 17106–17116, 2017.

[31] X. Deng, Y. Wen, and Y. Chen, "Highly efficient epidemic spreading model based lpa threshold community detection method," *Neurocomputing*, vol. 210, pp. 3–12, 2016.

[32] S. Fortunato and A. Lancichinetti, "Community detection algorithms: a comparative analysis: invited presentation, extended abstract," in *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, pp. 1-2, Pisa, Italy, October 2009.

[33] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding statistically significant communities in networks," *PLoS One*, vol. 6, no. 4, 2011.

[34] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.

[35] S. Chakrabarti, *Mining the Web: Analysis*, Morgan Kaufmann, Burlington, MA, USA, 2002.

[36] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowledge and Information Systems*, vol. 42, no. 1, pp. 181–213, 2015.

[37] J. Leskovec and J. J. Mcauley, "Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems*, pp. 539–547, Springer, Berlin, Germany, 2012.

Hindawi
Complexity
Volume 2020, Article ID 3917812, 10 pages
https://doi.org/10.1155/2020/3917812

Research Article

Joint Nonnegative Matrix Factorization Based on Sparse and Graph Laplacian Regularization for Clustering and Co-Differential Expression Genes Analysis

Ling-Yun Dai, Rong Zhu, and Juan Wang

School of Computer Science, Qufu Normal University, Rizhao 276826, China

Correspondence should be addressed to Ling-Yun Dai; dailingyun_1@163.com

Received 12 June 2020; Revised 8 October 2020; Accepted 12 October 2020; Published 16 November 2020

Academic Editor: Jia Wu

The explosion of multiomics data poses new challenges to existing data mining methods. Joint analysis of multiomics data can make the best of the complementary information that is provided by different types of data. Therefore, they can more accurately explore the biological mechanism of diseases. In this article, two forms of joint nonnegative matrix factorization based on the sparse and graph Laplacian regularization (SG-jNMF) method are proposed. In the method, the graph regularization constraint can preserve the local geometric structure of data. $L_{2,1}$-norm regularization can enhance the sparsity among the rows and remove redundant features in the data. First, SG-jNMF1 projects multiomics data into a common subspace and applies the multiomics fusion characteristic matrix to mine the important information closely related to diseases. Second, multiomics data of the same disease are mapped into the common sample space by SG-jNMF2, and the cluster structures are detected clearly. Experimental results show that SG-jNMF can achieve significant improvement in sample clustering compared with existing joint analysis frameworks. SG-jNMF also effectively integrates multiomics data to identify co-differentially expressed genes (Co-DEGs). SG-jNMF provides an efficient integrative analysis method for mining the biological information hidden in heterogeneous multiomics data.

1. Introduction

With the development of state-of-the-art sequencing technology, a large quantity of effective experimental data has been collected. These data may imply some unknown molecular mechanisms. Bioinformatics is faced with the task of analyzing massive omics data. The Cancer Gene Atlas (TCGA, https://tcgadata.nci.nih.gov/tcga/) includes gene expression profile data (GE), DNA methylation data (DM), copy number variation data (CNV), protein expression data, and drug sensitivity data. These data are from approximately 15,000 clinical samples of more than 30 kinds of cancers [1]. These massive data enable researchers to study the mechanisms of cancer production, diagnosis, and treatment at different biological levels.

The joint analysis of multiomics data can make up for lost or unreliable information in single omics data. In recent years, scientists have performed considerable research on the cancer mechanisms based on the joint analysis of cancer multiomics data. For example, Christina et al. integrated the gene expression data and copy number variations of breast cancer, identified possible pathogenic genes, and discovered new subtypes of breast cancer [2]. Wang and Wang used similarity network fusion to jointly analyze mRNA, DM, and microRNA (miRNA) data and identify cancer subtypes further [3]. In the existing joint analysis methods, those based on matrix decomposition are remarkable. Liu et al. integrated mRNA, somatic cell mutation, DNA methylation, and copy number variation data. They established a block constraint-based RPCA model to identify differentially expressed genes (DEGs) [4]. Integration and analysis of these heterogeneous multiomics data provide an in-depth understanding of the pathogenesis of cancer and promote the development of precision medicine. Recently, unsupervised integrative methods based on matrix decomposition have attracted considerable

attention among the existing methods for integrating and analyzing multiomics data. Zhang et al. constructed a joint matrix factorization framework (jNMF) to discover multidimensional modules of genomic data [5]. Yang and Michailidis introduced a new method named integrative NMF (iNMF) for heterogeneous multiomics data [6]. Strazar et al. incorporated orthogonality regularization into iNMF (iONMF) to integrate and analyze multiple data sources [7]. Joint nonnegative matrix decomposition meta-analysis (jNMFMA) [8], multiomics factor analysis (MOFA) [9], and Bayesian joint analysis [10] have been successfully applied to the integration and analysis of cancer omics data. To avoid the influence of redundant information, many sparse modeling methods have been proposed. Typical applications are as follows: The weighted sparse representation classifier (WSRC) model combined with global coding (GE) [11] was used to predict interactions between proteins based on protein sequence information. The network regularization sparse logic regression model (NSLR) [12] was used to predict survival risk and discover biomarkers. Sparse coregularization matrix decomposition was used to find mutant driver genes and so on [13].

In recent years, graph/network-based analysis as a powerful data representation tool has been applied to the modeling and analysis of complex systems [14–17]. In general, entities can be regarded as nodes, and the interaction between entities can be regarded as edges in the graph. Graph-based approaches can explore the local subspace structure and obtain the low-dimensional representation of high-dimensional data. Zhang and Ma proposed a subspace clustering algorithm based on a graph to detect the common modules highly correlated with cancer by jointly analyzing the gene expression and protein interaction networks [18]. Mixed-norm Laplacian regularized low-rank representation (MLLRR) was used to cluster samples [19]. Cui proposed an improved graph-based method to predict drug-target interactions [20]. Liu et al. introduced the contributions of deep neural networks, deep graph embedding, and graph neural networks along with the opportunities and challenges they faced [21]. Wu et al. proposed a multigraph learning algorithm called gMGFL that search and choose a group of decision subgraphs as features to move bags and bag labels to the instance [22].

Recently, sparse regularization has played a very important role in data analysis. The L_0-norm, L_1-norm, $L_{2,1}$-norm, etc. are all typical sparse regularization methods. Among these many sparse constraints, $L_{2,1}$-norm regularization stands out in terms of computational time and performance. The $L_{2,1}$-norm can obtain a sparse projection matrix in rows to learn discriminative features in the subspace. Zhang used the $L_{2,1}$-norm constraint on the coefficients to ensure that they are sparse in rows [23]. The $L_{2,1}$-norm was applied to the predictor to ensure that it is robust to noise and outliers [24].

Considering the role of graph regularizations and $L_{2,1}$-norm constraints in matrix factorization, we propose joint nonnegative matrix factorization based on sparse and graph Laplacian regularization (SG-jNMF). SG-jNMF can make the best of the potential associations and complementary information among multiomics data. The main highlights of this approach are as follows.

(1) Graph regularization is incorporated into the joint nonnegative matrix factorization model, and undirected graphs are constructed for input data in this method. Local graph regularization can preserve the local geometrical structure of the data space. Therefore, SG-jNMF can use the low-dimensional characteristics of the observed data to find intrinsic laws and improve the performance of the integrated analysis method.

(2) $L_{2,1}$-norm regularization can deal with each row of the matrix as a whole and can enhance the sparsity among the rows. Therefore, involving the $L_{2,1}$-norm can remove redundant features and noise in the data and further explore the clear cluster structure.

(3) Two forms of SG-jNMF are proposed. SG-jNMF1 projects multiomics data into a fusion feature space. The fusion matrix contains complementary and differential information provided by multiomics data, so that more accurate results can be obtained when identifying Co-DEGs. SG-jNMF2 projects multiomics data into a common sample space, which results in more accurate clustering results.

The rest of this paper is arranged as follows: In Section 2, we start with a brief review of jNMF. Next, we introduce the SG-jNMF method, optimization process, and computational complexity analysis. Section 3 gives out the experimental results of clustering and feature selection. Finally, we summarize the whole paper and give some suggestions for future work in Section 4.

2. Materials and Methods

2.1. Joint Nonnegative Matrix Factorization. The jNMF method was first proposed by Zhang et al. [5]. It can project multiple input data matrices into a common subspace, to integrate the information of each input data for analysis. Each type of genomic data as original data can be denoted as $X_I \in R^{M\times N}$ $(I = 1, 2, 3, \ldots)$. $W \in R^{M\times K}$ is the common basis matrix, and $H_I \in R^{K\times N}$ is the corresponding coefficient matrix. The objective function of jNMF can be written as

$$\begin{aligned} \min \quad & \sum_{I=1}^{P} \left\| X_I - WH_I \right\|_F^2, \\ \text{s.t.} \quad & W \geq 0, H_I \geq 0. \end{aligned} \tag{1}$$

Obviously, jNMF is the same as NMF when $P = 1$. Therefore, jNMF is the generalization model of NMF for multiple input datasets. Similar to NMF, multiplicative update rules are used to minimize the objective function. W and H_I are iteratively updated according to the following rules.

$$W_{ia} = W_{ia}\frac{\left(\sum_{I=1}^{P}\left(X_I H_I^T\right)\right)_{ia}}{\left(\sum_{I=1}^{P}\left(W\left(H_I H_I^T\right)\right)\right)_{ia}}, \tag{2}$$

$$H_{Iaj} = H_{Iaj}\frac{\left(W^T X_I\right)_{aj}}{\left(W^T W H_I\right)_{aj}}. \tag{3}$$

The jNMF method can be used to integrate and analyze multiomics data. It decomposes multiomics data matrices into multiple independent coefficient matrices and a common fusion matrix at the same time and projects high-dimensional omics data into low-dimensional spaces. Therefore, the abundant differential and complementary information of cancer multiomics data can be efficiently used, and multiomics datasets are analyzed simultaneously to obtain hidden information with biological significance.

2.2. Joint Nonnegative Matrix Factorization Based on Sparse and Graph Laplacian Regularization. Manifold learning has become a popular research topic in the domain of information science since it was first proposed in science in 2000 [25, 26]. Assuming that the data are uniformly sampled in a high-dimensional space, manifold learning can find the low-dimensional structure in the high-dimensional space and obtain the corresponding embedding mapping. Manifold learning looks for the essence of things from observed phenomena and finds the internal laws of data. The manifold assumption states that data points that are geometrically adjacent usually have similar characteristics. Therefore, an undirected weighted network/graph $G = (V; E; U)$ is constructed. $V = 1, 2, \ldots, P$ is the vertex set, E is the edge set, and U is the weight set. Edge weight $U_{j,k}(1 \le j \ne k \le q)$ is associated with edge(j, k) in E. The graph regularization with G is as follows:

$$\sum_k R_k = \sum_k \frac{1}{2}\sum_{i,j}^{N}\left(\left(f_k(x_i) - f_k(x_j)\right)^2 U_{i,j}\right) = \sum_k \frac{1}{2}\sum_{i,j}^{N}\left(\left(V_{i,j} - V_{k,j}\right)^2 U_{i,j}\right) = \sum_k V_k^T D V_k - \sum_k V_k^T U V_k = \sum_k V_k^T L V_k = T_r\left(V^T L V\right), \tag{4}$$

where $T_r(\cdot)$ is the trace of the matrix, L is the graph Laplacian matrix, and $L = D - U$. D is a diagonal matrix and $D_{i,j} = \sum_j U_{i,j}$. Intuitively, the smaller the R_k value is, the closer the two data points are. By minimizing R_k, we can obtain a sufficiently smooth mapping function on the data manifold.

To decrease the influence of noise and outliers on real data, sparse regularization is usually used to penalize the coefficient matrix. The L_0-norm, L_1-norm, and $L_{2,1}$-norm are all typical sparse regularization methods. The solution of L_0-norm is a NP-hard problem. L_1-norm is widely used because it has better optimization solution characteristics than L_0-norm. L_1-norm will tend to produce a small number of features, while the other features are all 0. Therefore, it can be used for feature selection. However, L_1-norm regularization is usually time-consuming. $L_{2,1}$-norm regularization on the coefficient matrix can generate a row sparse result, and the calculation of the $L_{2,1}$-norm is simple and convenient [23]. In this article, the $L_{2,1}$ penalty is incorporated in SG-jNMF [27]. The $L_{2,1}$-norm of a matrix Z is defined as

$$\|Z\|_{2,1} = \sum_{i=1}^{m}\sqrt{\sum_{j=1}^{n} z_{ij}^2} = \sum_{i=1}^{m}\left\|z^i\right\|_2. \tag{5}$$

2.2.1. SG-jNMF1. There are two forms of SG-jNMF methods in this article. As shown in Figure 1, the SG-jNMF1 method projects multiomics data into a common feature space. Graph regularization and a sparse penalty are applied to the fusion feature matrix. The feature matrix is constrained by graph regularization, and as much intrinsic geometric information of the original multiomics data are preserved as possible. The $L_{2,1}$-norm is used to constrain the feature matrix to reduce the influence of outliers and noise, and the objective function of integrating nonnegative matrix decomposition is constructed. The optimization problem can be expressed as

$$\begin{aligned} &\min \sum_{I=1}^{P}\left\|X_I - W H_I^T\right\|_F^2 + \sum_{I=1}^{P}\lambda_I T_r\left(W L_{I1} W^T\right) + \beta\|W\|_{2,1}, \\ &\text{s.t.} \quad W \ge 0, H_I \ge 0, \end{aligned} \tag{6}$$

where L_{I1} is the Laplacian matrix. $L_{I1} = D_{I1} - U_{I1}$, where U_{I1} is a symmetric matrix, which is the weight matrix constructed in graph regularization. D_{I1} is a diagonal matrix, and its diagonal elements are equal to the sum of the corresponding row elements or the sum of the column elements of the matrix; i.e., $D_{I1ii} = \sum_{j=1}^{n}(U_{I1ij})$.

With randomly positive initializing matrices $W$and H_I, the following update rules are executed until the algorithm converges:

$$\begin{aligned} W_{ia} &= W_{ia}\frac{\left(\sum_{I=1}^{P}\left(X_I H_I + \lambda_I U_{I1} W\right)\right)_{ia}}{\left(\sum_{I=1}^{P}\left(H_I H_I^T W + \lambda_I D_{I1} W\right) + \beta Q W\right)_{ia}}, \\ H_{Iaj} &= H_{Iaj}\frac{\left(X_I^T W^T\right)_{aj}}{\left(H_I W^T W\right)_{aj}}, \end{aligned} \tag{7}$$

where Q is a diagonal matrix, the diagonal element is $Q_{jj} = 1/\sqrt{\sum_{i=1}^{m}(W_{ij}) + \varepsilon}$, and ε is an infinitesimal positive number.

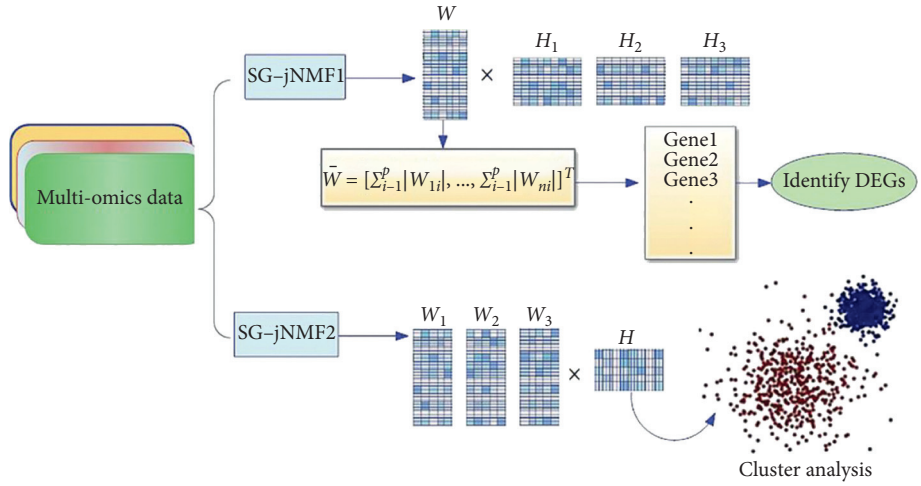


Figure 1: Framework of SG-jNMF.

2.2.2. SG-jNMF2. As seen from Figure 1, the SG-jNMF2 method projects multiomics data into a common sample space. Constraints are enforced on the common sample matrix. This method can be used to cluster multiomics data. The model can be shown by the following expression:

$$\min \sum_{I=1}^{P} \left\| X_I - W_I H^T \right\|_F^2 + \sum_{I=1}^{P} \lambda_I T_r\left(H^T L_{I2} H\right) + \beta \|H\|_{2,1},$$
$$\text{s.t.} \quad W_I \geq 0, H \geq 0. \tag{8}$$

Similarly, the algorithm iterates until it converges according to the following rules:

$$W_{Iia} = W_{Iia} \frac{(X_I H)_{ia}}{\left(W_I H^T H\right)_{ia}},$$
$$H_{aj} = H_{aj} \frac{\left(\sum_{I=1}^{P} \left(X_I^T W_I + \lambda_I U_{I2} H\right)\right)_{aj}}{\left(\sum_{I=1}^{P} \left(H W_I^T W + \lambda_I D_{I2} H\right) + \beta B H\right)_{aj}}, \tag{9}$$

where L_{I2} is the Laplacian matrix. $L_{I2} = D_{I2} - U_{I2}$, where $U_{I2}$is a symmetric matrix, which is the weight matrix constructed in graph regularization. D_{I2} is a diagonal matrix, and its diagonal elements are equal to the sum of the corresponding row elements or the sum of the column elements of the matrix; i.e., $D_{I2ii} = \sum_{j=1}^{n} (U_{I2ij})$. B is a diagonal matrix, and the diagonal element is $B_{jj} = 1/\sqrt{\sum_{i=1}^{m} (W_{ij}) + \varepsilon}$. Obviously, the objective functions of the two kinds of SG-jNMF method are both nonconvex. We can obtain the optimal solutions by minimizing the objective functions. The optimization process is shown as follows.

2.3. Optimization of SG-jNMF. Since the optimization processes of the two forms of SG-jNMF method are very similar, we only provide that of the first method. We use the multivariable alternating update rules to solve the optimization problem. Specifically, the following update steps are repeated until the algorithm converges.

2.3.1. Optimization of W. When H_I is fixed, the optimization of W is performed by minimizing the following objective function:

$$\begin{aligned} O &= \sum_{I=1}^{P} \left\| X_I - W H_I^T \right\|_F^2 + \sum_{I=1}^{P} \lambda_I T_r\left(W L_{I1} W^T\right) + \beta \|W\|_{2,1} \\ &= \sum_{I=1}^{P} \left(T_r\left(X_I^T X - 2 X_I^T W H_I^T + H_I W^T W H_I^T\right)\right) \\ &\quad + \lambda_I T_r\left(W L_{I1} W^T\right) + \beta T_r\left(W^T Q W\right). \end{aligned} \tag{10}$$

The corresponding Lagrangian function is as follows:

$$\begin{aligned} L = &\sum_{I=1}^{P} \left(T_r\left(X_I^T X - 2 X_I^T W H_I^T + H_I W^T W H_I^T\right)\right) \\ &+ \lambda_I T_r\left(W L_{I1} W^T\right) + \beta T_r\left(W^T Q W\right) \\ &+ T_r(\Phi W) + T_r\left(\Psi H_I^T\right), \end{aligned} \tag{11}$$

where $\Phi = [\phi_{il}]$ and $\Psi = [\psi_{Ia}]$ are the Lagrangian multipliers of $W$and H_I, respectively. Next, we take the first partial derivative of this Lagrangian function with respect to W:

$$\frac{\partial l}{\partial W} = \sum_{I=1}^{P} \left[-2 X H_I + 2 W H_I^T H_I + 2 \lambda_I L_{I1} W\right] + 2 \beta Q W + \Phi. \tag{12}$$

According to the KKT conditions [28], the following updating rule can be obtained:

$$W_{ia} = W_{ia} \frac{\left(\sum_{I=1}^{P} \left(X_I H_I + \lambda_I U_{I1} W\right)\right)_{ia}}{\left(\sum_{I=1}^{P} \left(H_I H_I^T W + \lambda_I D_{I1} W\right) + \beta Q W\right)_{ia}}. \tag{13}$$

2.3.2. Optimization of H_I. When W is fixed, the optimization of H_I is performed by minimizing the following objective function.

$$O = \sum_{I=1}^{P} \left\| X_I - W H_I^T \right\|_F^2, \quad \text{s.t.} \quad W \geq 0, H_I \geq 0. \tag{14}$$

The corresponding Lagrangian function is as follows:

$$l = \sum_{I=1}^{P} \left(T_r \left(X_I^T X - 2 X_I^T W H_I^T + H_I W^T W H_I^T \right) \right) + T_r(\Psi) H_I^T, \tag{15}$$

and H_I runs to convergence according to the following formula:

$$H_{Iaj} = H_{Iaj} \frac{\left(X_I^T W^T \right)_{aj}}{\left(H_I W^T W \right)_{aj}}. \tag{16}$$

2.4. Convergence and Running Time. In this paper, we also demonstrate the convergence of the method through experiments. Taking the pancreatic adenocarcinoma (PAAD) dataset as an example, the convergence of the five methods is shown in Figure 2. The error function used in this article is defined as follows:

$$\text{Loss} = \sum_{I=1}^{P} \frac{\left\| X_I - W H_I^T \right\|_F^2}{\left\| X_I \right\|_F^2}. \tag{17}$$

Compared with the other four methods, SG-jNMF can converge to the smallest error value with the fastest speed.

Besides, we also tested the running time of the above methods on the PAAD dataset. The means of these five methods running 10 times on a PC are shown in Table 1. As seen in Table 1, iGMFNA has the shortest running time, followed by SG-jNMF. This is due to the introduction of sparse constraints in SG-jNMF. The running time of iNMF, iGMFNA, jNMF, and SG-jNMF methods is satisfactory.

2.5. Computational Complexity Analysis. In this part, we discuss the extra computational complexity of SG-jNMF compared to jNMF. We use big O symbol to represent the computational complexity of the algorithm. On the basis of the updating rules (3) and (4), we can easily count the arithmetic operations of each iteration in jNMF. Obviously, the cost for each iteration in jNMF is $O(\text{MNk})$. It should be noted that U_I is a sparse matrix for SG-jNMF. In addition to the multiplicative updates, constructing a K-nearest neighbor graph requires $O(N^2M)$ operations [28]. Assume that the update stops after t iterations, and the overall cost for jNMF is $O(\text{tMNk})$. The overall cost for SG-jNMF is $O(N^2M) + O(\text{tPMNk})$.

3. Results and Discussion

3.1. Data Processing. TCGA project includes a lot of gene expression profile data, DNA methylation data, copy number variation data, protein expression data, drug sensitivity data, and so on. In-depth study of these data can help us to master the mechanism of cancer occurrence and development and provide technical support for prevention, diagnosis, and treatment of cancer. In this article, four cancer datasets which are all downloaded from TCGA (https://tcgadata.nci.nih.gov/tcga/), namely, PAAD, esophageal carcinoma (ESCA), cholangiocarcinoma (CHOL), and colon adenocarcinoma (COAD), are used in these experiments. Details are listed in Table 2. To avoid the matrix dimension problem in algorithm execution, the number of genes in the four datasets is aligned to 19,876. First, RPCA is used to reduce the effects of noise and redundant information [29]. Second, the same number of samples and characteristics is retained for multiomics data of the same kind of cancer. Then, the matrices are normalized according to the standard deviation of the data such that each element of the matrix is evaluated between 0 and 1.

3.2. Clustering. When SG-jNMF2 method projects multiomics data into a common sample space, it contains all the sample information provided by the input multiomics data. To assess the clustering performance of this method, SG-jNMF2 is used to cluster the tumor samples on CHOL, PAAD, COAD, and ESCA datasets. There are four methods (iNMF, iONMF, iGMFNA, and jNMF) that perform the same experiments on the same datasets.

3.2.1. Selection of Parameters. For SG-jNMF2, clustering performance is affected by the regularization parameters. In this experiment, we empirically set the same value for λ_I with different omics data from the same cancer [30]. Therefore, there are three parameters, λ, β, and K, that need to be adjusted. λ is the graph regularization parameter, β controls the sparsity of factorization, and K is the number of nodes in the undirected graph constructed in the manifold. From Figure 3, when K is set to 3, the accuracy on the four datasets reaches a maximum. As seen from Figure 4, λ should be set to 1,000 on PAAD. When λ is equal to 0.1, the accuracy on COAD can achieve the maximum. When λ is equal to 10^{-3}, 10^{-5}, and 1, the accuracy on CHOL can achieve the maximum. When λ is equal to 10^4, the accuracy on ESCA can achieve the maximum. From Figure 5, when β is set from 10^{-5} to 10^1 for PAAD, the accuracy reaches the maximum. For ESCA and COAD, β should be set from 10^4 to 10^5. For CHOL, the value of β does not matter much.

3.2.2. Evaluation Indicators. Several indicators are used to evaluate the clustering performance of SG-jNMF2: accuracy, recall, precision, and F1-score. Accuracy is defined as

$$\text{AC} = \frac{\sum_{j=1}^{N} \delta\left(s_j, \text{map}\left(r_j \right) \right)}{N}, \tag{18}$$

where N is the total number of samples in the dataset and $\delta(x, y)$ is a singular function. When x is equal to y, the value of the function is equal to 1; otherwise, it is equal to 0. $\text{map}(r_j)$ maps the clustering label r_j to the real label s_j. The other three indicators used to evaluate clustering performance are defined as follows:

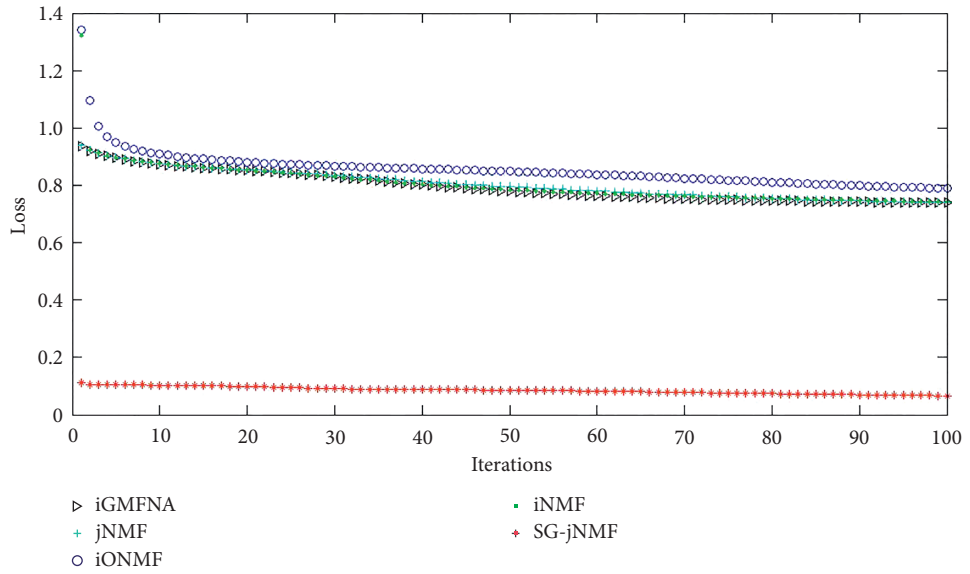


Figure 2: Comparison of convergence of five methods on PAAD dataset.

Table 1: Running time on PAAD.

Methods	Running times
iNMF	2.4261
iGMFNA	0.1312
jNMF	1.6311
iONMF	46.8843
SG-jNMF	0.5153

Table 2: Overview of multiomics datasets.

Multiomics datasets	Total number of samples	Cancer samples	Number of genes
PAAD (GE, ME, CNV)	180	176	19877
CHOL (GE, ME, CNV)	45	36	19876
ESCA (GE, ME, CNV)	192	183	19877
COAD (GE, ME, CNV)	281	262	22723

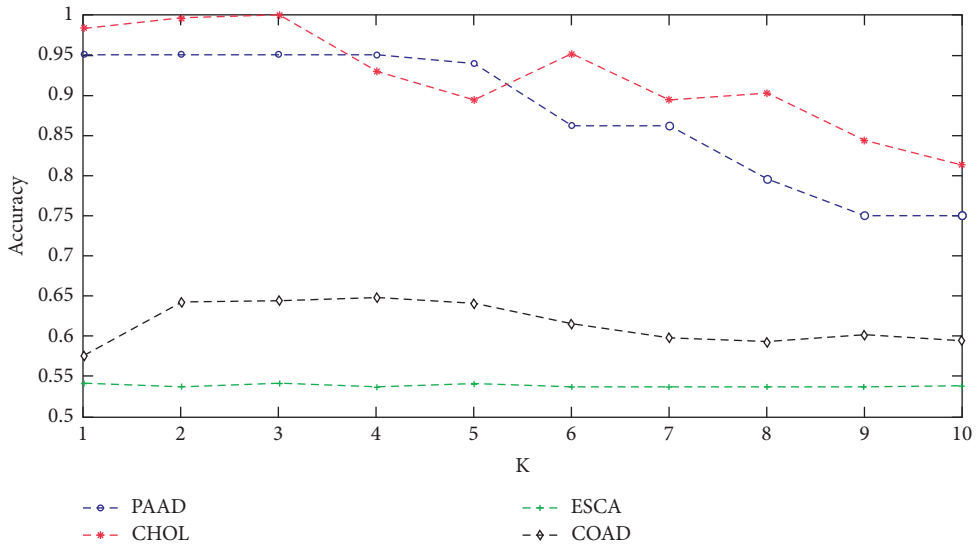


Figure 3: Accuracy of SG-jNMF varies with K.

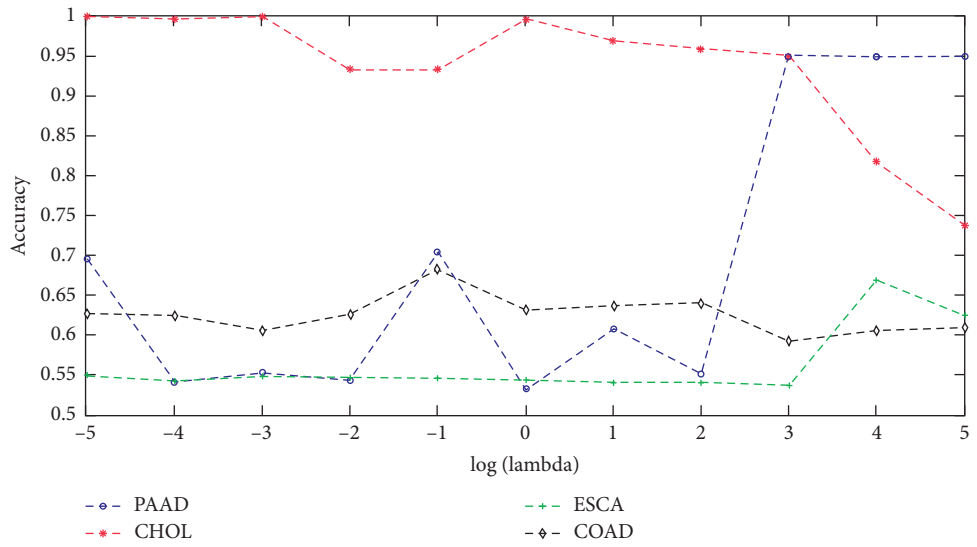


Figure 4: Accuracy of SG-jNMF varies with λ.

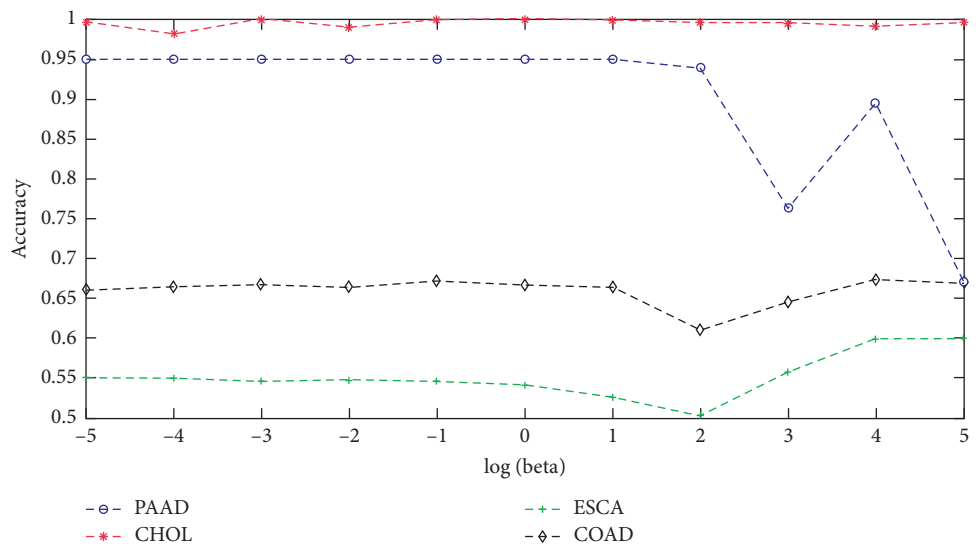


Figure 5: Accuracy of SG-jNMF varies with β.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$
$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{19}$$
$$F1 - \text{score} = \frac{2}{(1/\text{recall}) + (1/\text{precision})},$$

where TP means the number of true positives, FP is the number of false positives, and FN denotes the number of false negatives.

3.2.3. Results. In this experiment, each algorithm was run fifty times to reduce the impact of random initialization on the clustering results. We compared the accuracy, recall, precision, and F1-score of the four methods with SG-jNMF2. The mean and variance in the results are shown in Table 3. As seen in Table 3, SG-jNMF2 achieves the highest values on the four indicators mentioned above, except the recall value on the ESCA dataset. The contributions of sparse and graph regularization constraints of the algorithm are listed in Table 4. Performance improvements are measured by $\Delta_{\text{ind}} = (\text{Ind}_i - \text{Ind}_j)/(\text{Ind}_j)$, where Ind_i is the indicator of SG-jNMF and Ind_j is that of the comparison method. In particular, sparse constraints improve accuracy by 49.70%, and sparse and graph regularization constraints improve accuracy by 78.87% on the PAAD dataset. Recall and F1-score achieve more than 50% improvement on the CHOL dataset. When sparse constraints are introduced, only the recall on ESCA is reduced by 0.53%. The results on other datasets have also improved to varying degrees. In summary,

Table 3: Performance of different analysis methods.

	Datasets	iNMF	iGMFNA	jNMF	iONMF	SG-jNMF
Accuracy	PAAD	53.56 (0.00)	63.44 (3.05)	53.11 (0.01)	56.23 (0.72)	**95.00 (0.00)**
	CHOL	90.22 (1.71)	97.33 (0.04)	93.78 (0.49)	90.04 (0.96)	**99.11 (0.00)**
	ESCA	54.17 (0.00)	54.58 (0.00)	53.96 (0.00)	58.70 (1.46)	**66.87 (0.01)**
	COAD	61.87 (0.01)	63.35 (0.02)	59.15 (0.09)	62.13 (0.22)	**68.84 (0.01)**
Recall	PAAD	51.47 (4.68)	58.07 (4.84)	53.34 (4.50)	48.44 (1.20)	**67.33 (2.26)**
	CHOL	50.78 (2.01)	56.17 (1.84)	55.39 (2.04)	46.41 (1.49)	**88.06 (0.47)**
	ESCA	50.98 (0.29)	**51.31 (0.29)**	48.86 (0.27)	51.13 (0.30)	51.04 (0.12)
	COAD	49.34 (1.25)	49.15 (1.30)	49.65 (1.28)	46.73 (1.08)	**56.18 (1.51)**
Precision	PAAD	97.79 (0.05)	98.60 (0.04)	97.52 (0.05)	97.71 (0.05)	**99.09 (0.01)**
	CHOL	58.55 (0.20)	62.21 (0.20)	63.36 (0.21)	60.54 (1.72)	**91.00 (0.10)**
	ESCA	94.40 (0.05)	95.20 (0.07)	95.90 (0.08)	95.67 (0.06)	**98.39 (0.02)**
	COAD	49.34 (1.25)	49.15 (1.30)	49.65 (1.28)	46.73 (1.08)	**56.18 (0.51)**
F1-score	PAAD	63.54 (3.70)	66.70 (1.29)	67.84 (1.71)	64.42 (1.86)	**77.15 (2.12)**
	CHOL	53.47 (2.01)	58.19 (1.91)	58.03 (2.04)	51.61 (1.63)	**89.18 (1.04)**
	ESCA	66.00 (0.18)	66.47 (0.19)	64.52 (0.18)	66.41 (0.16)	**66.96 (0.16)**
	COAD	63.71 (1.34)	63.56 (1.27)	64.25 (1.30)	61.23 (1.14)	**71.11 (0.55)**

Table 4: The contribution of graph regularization and sparse constraints to clustering performance.

	Datasets	Beta = 0 (%)	Lambda = 0, beta = 0 (%)
Accuracy	PAAD	49.70	78.87
	CHOL	1.82	5.68
	ESCA	22.52	23.93
	COAD	8.67	16.38
Recall	PAAD	15.90	26.22
	CHOL	56.77	58.98
	ESCA	−0.53	4.46
	COAD	14.30	13.15
Precision	PAAD	0.49	1.60
	CHOL	46.28	43.62
	ESCA	3.35	2.59
	COAD	6.18	4.92
F1-score	PAAD	15.67	13.72
	CHOL	53.26	53.68
	ESCA	0.70	3.78
	COAD	11.88	10.68

the performance of the integrated NMF in analyzing multiomics data greatly improves by introducing sparse constraints and graph regularization constraints.

3.3. Identifying Co-DEGs. First, three matrices (DM, GE, and CNV of PAAD) are input into the SG-jNMF1 model and are projected into a common feature space. Second, we sum the common feature matrix in rows. Finally, we sort the elements in the sum vector in descending order. The top 100 genes are selected as Co-DEGs. These 100 genes are compared with pancreatic cancer genes exported from GeneCards (URL:http://www.genecards.org). Co-DEGs with relevance scores above 4 are listed in Table 5. CDKN2A is frequently mutated or deleted in many tumors. It plays an important role as a tumor suppressor gene. Studies have shown that the mutation of CDKN2A is closely related to the development of pancreatic cancer in families [31]. It is frequently seen in many tumors that mutation and overexpression of CCDN1 can alter the process of the cell cycle. Wang et al. identified pancreatitis-associated genes and found that CCND1 was involved in the pathway of pancreatic cancer [32]. Research on transcriptome sequencing shows that PTF1A maintains the expression of genes in all cellular processes. Deletion of PTF1A leads to an imbalance, cell damage, and acinar metaplasia, which is directly related to the development of pancreatic cancer [33]. Scientists have explored the effects of GRP on human intestinal and pancreatic peptides. Therefore, SG-jNMF1 can effectively integrate the information of multiomics data to identify Co-DEGs closely related to the disease.

We also use SG-jNMF1 to integrate three gene expression datasets from ESCA, CHOL, and COAD to identify Co-DEGs associated with all three diseases. Partially Co-DEGs and their relevance scores with ESCA, CHOL, and COAD are shown in Table 6. The relevance score of CHEK2 with ESCA is up to 77.66. Allelic variation in CHEK2 has a strong relationship with the risk of esophageal cancer [34]. Relevance score of CHEK2 with COAD is 29.65. The germline variation in CHEK2 is also closely related to the

Table 5: Co-DEGs identified by SG-jNMF on PAAD.

Name	Relevance score	Associated diseases	Related pathways
CDKN2A	91.19	Melanoma, cutaneous malignant 2, and melanoma-pancreatic cancer syndrome	Modulation and signaling and cell cycle role of SCF complex in cell cycle regulation
CCDN1	53.49	Multiple myeloma and Von Hippel–Lindau syndrome	ATF-2 transcription factor network and DNA damage response
PTF1A	33.66	Pancreatic and cerebellar agenesis and pancreatic agenesis 2	Developmental biology and regulation of beta-cell development
GRP	21.84	Duodenal ulcer and lung disease	Peptide ligand-binding receptors and signaling by GPCR

Table 6: Co-DEGs identified by SG-jNMF of ESCA, CHOL, and COAD.

Genes	Related diseases	Relevance score with ESCA	Relevance score with CHOL	Relevance score with COAD
CHEK2	Prostate cancer and Li–Fraumeni syndrome 2	77.66	0.35	29.65
BRAF	Cardiofaciocutaneous syndrome 1 and lung cancer	55.71	13.44	67.81
RARB	Microphthalmia, syndromic 12 and chromosome 3P deletion	28.39	2.5	23.41
NFE2L2	Immunodeficiency, developmental delay, and hypohomocysteinemia and lung squamous cell carcinoma	25.81	2.54	28.18

risk of colorectal cancer [35]. Frequent mutations in BRPA have been widely reported in human malignancies, including esophageal cancer, cholangiocarcinoma, and colon cancer [36–38]. This provides a computational method for the study of Co-DEGs in multiple diseases.

4. Conclusions

In this paper, we propose an integrative matrix factorization method (SG-jNMF) used to analyze heterogeneous multiomics data. The novel method jointly projects multiomics data matrices into a common low-dimensional space. Two forms of SG-jNMF enable multiomics data to be analyzed from both the sample and feature perspectives. This integrative analysis method can consider the local association of data and decrease the interference of noise and redundant information in the heterogeneous multiomics data. Experimental results show that the new method is superior to existing methods in analyzing heterogeneous multiomics data. Another significant advantage of SG-jNMF is that it can flexibly handle multiple input data of various types. This flexibility means that the input data can be different types of data (GE, ME, CNV, etc.) for the same disease or the same type of data for different diseases. We can use this method to identify Co-DEGs associated with a particular disease and detect common Co-DEGs associated with several diseases. This provides an efficient calculation method for biological and medical research. Next, we will use the correlation between Co-DEGs to build a gene coexpression correlation network, and further study the function of gene modules and related pathways.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the grants from the National Natural Science Foundation of China, nos. 61902215 and 61702299.

References

[1] J. N. Weinstein, E. A. Collisson, G. B. Mills et al., "The cancer genome atlas pan-cancer analysis project," *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, 2013.

[2] C. Curtis, S. P. Shah, S.-F. Chin et al., "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups," *Nature*, vol. 486, no. 7403, pp. 346–352, 2012.

[3] Y. L. Wang and X. Wang, "Fault diagnosis of wind turbine's converter based on memristive neural network," *Applied Mechanics and Materials*, vol. 705, no. 3, pp. 333–337, 2014.

[4] J.-X. Liu, Y.-L. Gao, C.-H. Zheng, Y. Xu, and J. Yu, "Block-constraint robust principal component analysis and its application to integrated analysis of tcga data," *IEEE Transactions on Nanobioscience*, vol. 15, no. 6, pp. 510–516, 2016.

[5] S. Zhang, C.-C. Liu, W. Li, H. Shen, P. W. Laird, and X. J. Zhou, "Discovery of multi-dimensional modules by integrative analysis of cancer genomic data," *Nucleic Acids Research*, vol. 40, no. 19, pp. 9379–9391, 2012.

[6] Z. Yang and G. Michailidis, "A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data," *Bioinformatics*, vol. 30, no. 1, pp. 1–8, 2015.

[7] M. Strazar, M. Zitnik, B. Zupan, J. Ule, and T. Curk, "Orthogonal matrix factorization enables integrative analysis of multiple RNA binding proteins," *Bioinformatics*, vol. 32, no. 10, pp. 1527–1535, 2016.

[8] H.-Q. Wang, C.-H. Zheng, and X.-M. Zhao, "jnmfma: a joint non-negative matrix factorization meta-analysis of

transcriptomics data," *Bioinformatics*, vol. 31, no. 4, pp. 572–580, 2014.

[9] R. Argelaguet, B. Velten, D. Arnol et al., "Multi-omics factor analysisâĂŤa framework for unsupervised integration of multi-omics data sets," *Molecular Systems Biology*, vol. 14, no. 6, pp. 1–13, 2018.

[10] P. Ray, L. Zheng, J. Lucas, and L. Carin, "Bayesian joint analysis of heterogeneous genomics data," *Bioinformatics*, vol. 30, no. 10, pp. 1370–1376, 2014.

[11] Y.-A. Huang, Z.-H. You, X. Chen, K. Chan, and X. Luo, "Sequence-based prediction of protein-protein interactions using weighted sparse representation model combined with global encoding," *BMC Bioinformatics*, vol. 171, no. 1, p. 184, 2016.

[12] W. Min, J. Liu, and S. Zhang, "Network-regularized sparse logistic regression models for clinical risk prediction and biomarker discovery," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 3, pp. 944–953, 2018.

[13] J. Xi, M. Wang, and A. Li, "Discovering mutated driver genes through a robust and sparse co-regularized matrix factorization framework with prior information from mirna expression patterns and interaction network," *BMC Bioinformatics*, vol. 19, no. 1, p. 214, 2018.

[14] Y. Pei, N. Chakraborty, and K. Sycara, "Nonnegative matrix tri-factorization with graph regularization for community detection in social networks," *ACM Transactions on Intelligent Systems and Technology, (TIST)*, vol. 8, no. 1, p. 1, 2016.

[15] L.-Y. Dai, C.-H. Zheng, J.-X. Liu et al., "Integrative graph regularized matrix factorization for drug-pathway associations analysis," *Computational Biology and Chemistry*, vol. 78, pp. 474–480, 2019.

[16] K. Zhan, J. Shi, J. Wang, and F. Tian, "Graph-regularized concept factorization for multi-view document clustering," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 411–418, 2017.

[17] J. Wu, S. Pan, X. Zhu et al., "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.

[18] E. Zhang and X. Ma, "Regularized multi-view subspace clustering for common modules across cancer stages," *Molecules*, vol. 23, no. 5, p. 1016, 2018.

[19] J. Wang, J.-X. Liu, C.-H. Zheng, Y.-X. Wang, X.-Z. Kong, and C.-G. Wen, "A mixed-norm Laplacian regularized low-rank representation method for tumor samples clustering," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 1, pp. 172–182, 2019.

[20] Z. Cui, "L 2, 1-grmf: an improved graph regularized matrix factorization method to predict drug-target interactions," *BMC Bioinformatics*, vol. 20, no. 8, p. 287, 2019.

[21] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," 2020, https://arxiv.org/abs/2005.08225.

[22] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[23] Z. Zhang, "Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3798–3814, 2017.

[24] Z. Zhang, Y. Zhang, S. Li et al., "Flexible auto-weighted local-coordinate concept factorization: a robust framework for unsupervised clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 1, p. 1, 2019.

[25] L.-Y. Dai, C.-M. Feng, J.-X. Liu, C.-H. Zheng, J. Yu, and M.-X. Hou, "Robust nonnegative matrix factorization via joint graph laplacian and discriminative information for identifying differentially expressed genes," *Complexity*, vol. 2017, Article ID 4216797, 11 pages, 2017.

[26] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5550, pp. 2268-2269, 2000.

[27] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint l2, 1-norms minimization," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 1813–1821, British Columbia, Canada, December 2010.

[28] D. Cai, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2010.

[29] M.-X. Hou, Y.-L. Gao, J.-X. Liu, L.-Y. Dai, X.-Z. Kong, and J. Shang, "Network analysis based on low-rank method for mining information on integrated data of multi-cancers," *Computational Biology and Chemistry*, vol. 78, pp. 468–473, 2019.

[30] N. Yu, Y.-L. Gao, J.-X. Liu, J. Shang, R. Zhu, and L.-Y. Dai, "Co-differential gene selection and clustering based on graph regularized multi-view NMF in cancer genomic data," *Genes*, vol. 9, no. 12, p. 586, 2018.

[31] D. B. Zhen, K. G. Rabe, S. Gallinger et al., "Brca1, brca2, palb2, and cdkn2a mutations in familial pancreatic cancer: a pacgene study," *Genetics in Medicine*, vol. 17, no. 7, p. 569, 2015.

[32] D. Wang, Z.-M. Zhu, Y.-L. Tu et al., "Identfication of key miRNAs in pancreatitis using bioinformatics analysis of microarray data," *Molecular Medicine Reports*, vol. 14, no. 6, pp. 5451–5460, 2016.

[33] C. Q. Hoang, M. A. Hale, Ana C. Azevedo-Pouly et al., "Transcriptional maintenance of pancreatic acinar identity, dierentiation, and homeostasis by ptf1a," *Molecular and Cellular Biology*, vol. 36, no. 34, pp. 3033–3047, 2016.

[34] H. Gu, W. Qiu, Y. Wan et al., "Variant allele of CHEK2 is associated with a decreased risk of esophageal cancer lymph node metastasis in a Chinese population," *Molecular Biology Reports*, vol. 39, no. 5, pp. 5977–5984, 2012.

[35] L. H. Williams, D. Choong, S. A. Johnson, and I. G. Campbell, "Genetic and epigenetic analysis of CHEK2 in sporadic breast, colon, and ovarian cancers," *Clinical Cancer Research*, vol. 12, no. 23, pp. 6967–6972, 2006.

[36] C. H. Maeng, J. Lee, P. van Hummelen et al., "High-throughput genotyping in metastatic esophageal squamous cell carcinoma identifies phosphoinositide-3-kinase and BRAF mutations," *PloS One*, vol. 7, no. 8, Article ID e41655, 2012.

[37] S. Robertson, O. Hyder, R. Dodson et al., "The frequency of KRAS and BRAF mutations in intrahepatic cholangiocarcinomas and their correlation with clinical outcome," *Human Pathology*, vol. 44, no. 12, pp. 2768–2773, 2013.

[38] S. Ogino, K. Nosho, G. J. Kirkner et al., "CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer," *Gut*, vol. 58, no. 1, pp. 90–96, 2009.

Hindawi
Complexity
Volume 2020, Article ID 8828087, 13 pages
https://doi.org/10.1155/2020/8828087

Research Article

Deep Interest-Shifting Network with Meta-Embeddings for Fresh Item Recommendation

Zhao Li,[1] **Haobo Wang**,[2] **Donghui Ding**,[1] **Shichang Hu**,[1] **Zhen Zhang**,[2] **Weiwei Liu**,[3] **Jianliang Gao**,[4] **Zhiqiang Zhang**,[5] **and Ji Zhang**[6]

[1]*Alibaba Group, Hangzhou, China*
[2]*Zhejiang University, Hangzhou, China*
[3]*Wuhan University, Wuhan, China*
[4]*Central South University, Changsha, China*
[5]*Zhejiang University of Finance and Economics, Hangzhou, China*
[6]*University of Southern Queensland, Toowoomba, Queensland, Australia*

Correspondence should be addressed to Jianliang Gao; gaojianliang@csu.edu.cn

Received 18 August 2020; Revised 22 September 2020; Accepted 8 October 2020; Published 28 October 2020

Academic Editor: Jia Wu

Nowadays, people have an increasing interest in fresh products such as new shoes and cosmetics. To this end, an E-commerce platform Taobao launched a fresh-item hub page on the recommender system, with which customers can freely and exclusively explore and purchase fresh items, namely, the *New Tendency* page. In this work, we make a first attempt to tackle the fresh-item recommendation task with two major challenges. First, a fresh-item recommendation scenario usually faces the challenge that the training data are highly deficient due to low page views. In this paper, we propose a deep interest-shifting network (DisNet), which transfers knowledge from a huge number of auxiliary data and then shifts user interests with contextual information. Furthermore, three interpretable interest-shifting operators are introduced. Second, since the items are fresh, many of them have never been exposed to users, leading to a severe cold-start problem. Though this problem can be alleviated by knowledge transfer, we further babysit these fully cold-start items by a relational meta-Id-embedding generator (RM-IdEG). Specifically, it trains the item id embeddings in a learning-to-learn manner and integrates relational information for better embedding performance. We conducted comprehensive experiments on both synthetic datasets as well as a real-world dataset. Both DisNet and RM-IdEG significantly outperform state-of-the-art approaches, respectively. Empirical results clearly verify the effectiveness of the proposed techniques, which are arguably promising and scalable in real-world applications.

1. Introduction

E-commerce has been prevalent in our daily life. In traditional online shopping scenarios, all items are mixed up, and a recommender system predicts users' preferences on items based on their past interactions, e.g., click, purchase, and rating [1–3]. However, this strategy overlooks the influence of the items' life periods and causes two problems. First, as many people have a growing interest in novel, newly released commodities, their requirements will not be fully satisfied. Second, popular items have more opportunities to be exposed, whereas those new products are overwhelmed, even though with high quality [4–6].

To tackle these problems, one E-commerce platform Taobao launched a new application, namely, *New Tendency* page, aiming to recommend fresh items for users who prefer new products. As illustrated in Figure 1, a card which contains a fresh item with its textual descriptions is pushed to the users. Once a user clicks this card, the *New Tendency* page appears, where more items from a predefined fresh item pool are recommended to this user. As a result, users who prefer newly released products can freely explore this

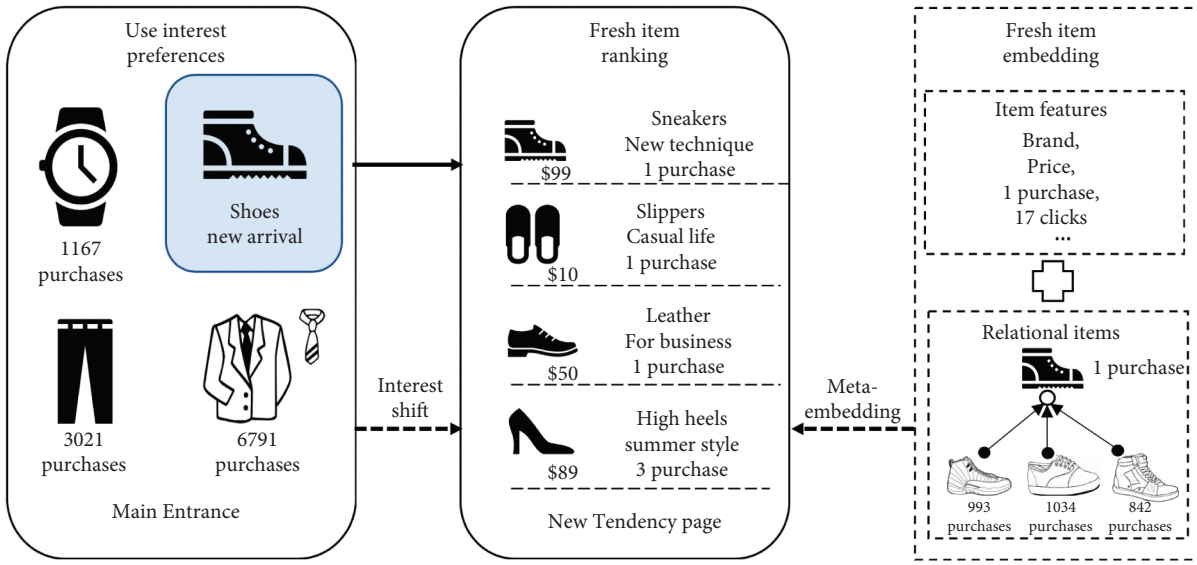


Figure 1: An illustration of fresh item recommended. Once a user clicks the card on the left side, a fresh-item recommendation page (middle side) appears to achieve interest shifting from users' history interests to the fresh items, where the recommended fresh items are ranked using the embeddings of fresh items and relational items (right side). All the fresh items on the recommended page are chosen from a specific fresh item pool, which are usually less exposed than items on the main entrance page.

page. However, to achieve high-quality ranking on this page, two key problems have to be addressed.

1.1. Q1: How to Address the Data-Deficiency Problem? Recommending fresh items directly on the main entrance page of the app may cause unpredictable influences. Thus, this page has to be designed as a fresh-item recommendation scenario. Compared to the main entrance page of the app, the *New Tendency* page is reported to contain less than 5% of page views. Most of the fresh items only have a few interactions, which make the scenario-specific training data highly deficient. As a result, we have to collect additional information to improve the performance.

1.2. Potential Solutions to Q1. We firstly notice that the clicked card contains rich contextual information, such as the showpiece and its textual description, which clearly reflects the user interests. Therefore, we can utilize off-the-shelf context-aware recommender systems (CARSs) [7], such as factorization-based approaches [8, 9] and deep learning-based models [10–15]. However, the model complexity increases owing to the involvement of the context features, which prevent the model being trained sufficiently. To deal with this problem, cross-domain recommender systems (CDRSs) [16–18] seem appealing due to their superiority in handling data deficiency. In particular, an asymmetric CDRS [19–21], which collects a large amount of context-free data (e.g., data from the main entrance of the app, namely, auxiliary data), can be designed to improve the prediction performance. However, existing asymmetric CDRS models seldom consider the scenario-specific contextual information of the target domain.

1.3. Q2: How to Deal with Totally Cold-Start Items? As reported by Taobao, more than 60% of fresh items are newborn and never interacted by users, which causes a severe cold-start problem. Note that these newborn items are not the cause of data deficiency because they are not a part of training data.

1.4. Potential Solutions to Q2. The cold-start problem is usually solved by integrating external information, e.g., item attributes [22, 23], user attributes [24, 25], relational data [26], and knowledge from other domains [16]. We note that this problem can be alleviated by applying the cross-domain technique because the embeddings of item attributes can be reused. Nevertheless, since the id of a cold-start item never appears, its embedding cannot obtain a good initialization. Pan et al. [27] proposed the meta-Id-embedding generator (Meta-IdEG), which considers the id embedding initialization problem and solves it through a learning-to-learn training manner. However, meta-IdEG only utilizes item features to generate the id embedding. As a result, it is unable to explore the community structural information when initializing id embeddings, which leads to a suboptimal solution.

1.5. Our Solutions. In this study, we propose two novel techniques to construct a deep learning-based recommender system, which simultaneously tackles these above issues. The proposed model fully exploits various types of external information to improve the prediction performance. To answer Q1, we present a deep interest-shifting network (DisNet). Specifically, it firstly learns the users' general interest vectors using a huge number of auxiliary data and then shifts them to a scenario-specific representation using

contexts. Next, the size of trainable parameters is reduced to a few neural network layers, which significantly alleviates the data-deficiency problem. To answer Q2, the transferred embedding layer of item attributes can be reused, and the only thing that matters is the item id embedding initialization problem. Hence, this paper proposes a relational meta-Id-embedding generator (RM-IdEG), which is trained in a learning-to-learn manner, aiming to make the model achieve great generalization ability after few-shot training. Furthermore, RM-IdEG absorbs the information of relevant items. Therefore, the community structural information can be inherently embedded and exploited, which has been proved beneficial for addressing the cold-start problem [26].

The main contributions of this work are summarized as follows:

A novel application, fresh item recommendation, is studied, which gives new items more opportunities to be exposed and fully personalizes the recommendations of those who prefer the novel, innovative products. We also make a first attempt to address the fresh-item recommendation task by two novel techniques.

We present a deep interest-shifting network (DisNet) to deal with the severe data-deficiency problem in a fresh-item recommendation scenario.

To address the cold-start problem, we propose a relational meta-Id-embedding generator (RM-IdEG) that involves the relational data into meta-id embedding initialization, which enables community structural information to be inherently contained.

Extensive experimental results demonstrate that our model can effectively handle fresh-item recommendation tasks in both cold-start and warm-start stages.

The rest of this work is organized as follows. In the next section, notations and preliminary knowledge are introduced. In Section 3, we provide a detailed description of our network architecture. After that, the results of empirical studies are reported. Then, we give the related works of our method. Further discussion and concluding remarks are provided in the last section.

2. Notations and Preliminaries

In this section, we firstly discuss a popular architecture of context-aware recommender systems. Then, we introduce the training procedure of meta-IdEG and summarize the notations in Table 1.

2.1. Context-Aware Recommendation. A popular strategy in existing context-aware recommendation systems is to learn latent representations for users and items and then make decisions using these latent vectors.

Formally, given an example, which contains an item t, a user u, and potentially some contexts, we first feed them into an embedding layer. Then, their features are transformed into vector representations by one-hot encoding or multihot encoding. The transformed item features consist of an item id embedding $\mathbf{e}_t$ and other content features $\mathbf{v}_t$. For the user, we combine its id embedding and other features as one vectorized representation $\mathbf{v}_u$. Finally, we denote the transformed context features by $\mathbf{c}$. The final prediction is made by

$$\begin{aligned} \widehat{y} &= g(\mathbf{q}_\mathbf{u}, \mathbf{p}_t, c), \\ \mathbf{q}_\mathbf{u} &= f_\mathbf{u}(\mathbf{v}_\mathbf{u}), \\ \mathbf{p}_t &= f_t(\mathbf{e}_t, \mathbf{v}_t). \end{aligned} \tag{1}$$

For example, in matrix factorization-based models [28], $\mathbf{q}_u$ and $\mathbf{p}_t$ are exactly their id embeddings, and g is the context-biased prediction function. State-of-the-art models [29, 30] also use neural networks to learn user/item representation as well as make decisions. This paper also adopts neural networks for f_t, f_u, and g, which lead to a double-tower model architecture.

It is noteworthy that such a learning paradigm deeply couples the contextual information in the model architecture. In our cross-domain setting, there are heterogeneous contexts, i.e., scenario-specific contexts. Therefore, the trainable parameters of deep neural network models cannot be reused, which makes them hard to share knowledge across different domains [31–33].

2.2. Meta-Id Embedding Generator. To babysit newborn items, the only thing that matters is how to learn the embeddings for new items' ids. A common learning paradigm first uses an Id embedding generator (IdEG) to initialize a vector for new ids in the embedding table and then update them using incoming user interactions. The most intuitive way is to output a random embedding initialization. However, its generalization ability may be restricted due to the cold-start problem. To this end, Pan et al. [27] proposed to initialize id embeddings using meta-learning technique, a.k.a. meta-Id embedding generator (meta-IdEG). By regarding the recommendation for each item as a task, meta-IdEG ensures good embedding initialization such that the model achieves better generalization ability after few-shot training.

Next, we illustrate the workflow of meta-IdEG. For each task that relates to a specific item, we divide its data examples (interactions) into two sets: a support set $\mathscr{D}^s$ and a query set $\mathscr{D}^q$. We firstly feed the item features into a neural network to generate an id embedding, $\mathbf{e}_t^* = \mathrm{IdEG}_{\mathrm{meta}}(\mathbf{v}_t)$. Then, we optimize $\mathrm{IdEG}_{\mathrm{meta}}$ in a learning-to-learn manner. We denote the predicted label on the support set as $\widehat{y}^*$ using $\mathbf{e}_t^*$. First, we can obtain the cold-start loss by

$$l_c(y, \widehat{y}^*; \mathbf{e}_t^*, \mathscr{D}^s) = -y\log(\widehat{y}^*) - (1-y)\log(1-\widehat{y}^*). \tag{2}$$

Then, we update the embedding by one step of gradient descent:

$$\mathbf{e}_t' = \mathbf{e}_t^* - \alpha\frac{\partial l_c}{\partial \mathbf{e}_t^*}, \tag{3}$$

where α is the learning rate. Since a new embedding is obtained, we can predict label $\widehat{y}'$ on the query set using $\mathbf{e}_t'$. Next, we define a warmed loss by

$$l_w(y, \widehat{y}'; \mathbf{e}_t', \mathscr{D}^q) = -y\log(\widehat{y}') - (1-y)\log(1-\widehat{y}'). \tag{4}$$

Table 1: Notations.

Notation	Definition or descriptions
(u, t)	A pair of user u and item t
$\mathbf{e}_t$	The id embedding of item t
$\mathbf{v}_t$	The embedded features (except id) of item t
$\mathbf{v}_u$	The embedded features (include id) of user u
$\mathbf{c}$, $\mathbf{c}_s$	Common contexts and scenario-specific contexts
f_t, f_u	Item/user representation networks
g	Decision-making network
h	Context network
$\mathbf{q}_u$, $\mathbf{p}_t$	Latent user/item vectors in the interest space
$\mathbf{z}_s$, $\mathbf{Z}_s$	Output vector/matrix of context network h
$\mathbf{e}_t^*$, $\mathbf{e}_t^{'}$	Cold-start and warmed item id embeddings in the meta-training procedure
$\mathscr{D}^s$, $\mathscr{D}^q$	Support and query sets of a cold-start item in $\mathscr{D}^c$
$\widehat{y}^*$, $\widehat{y}'$	Predicted labels on $\mathscr{D}^s$ and $\mathscr{D}^q$
l_c, l_w, l_{meta}	Cold-start, warmed, and meta-loss in the meta-training procedure
$\mathscr{D}^a$, $\mathscr{D}^w$, $\mathscr{D}^c$	Auxiliary, warm-start, and cold-start datasets
y, $\widehat{y}$, $\widehat{y}^{\text{DN}}$	Ground-truth label and predicted labels of context-aware models and DisNet
$\text{ISO}(\cdot)$	Interest-shifting operator
$\mathbf{e}_t^i$ $(i = 1, \ldots, k)$	The id embeddings of top k relevant items
$\mathbf{a}_t^i$ $(i = 1, \ldots, k)$	Attention scores
$\mathbf{W}_i$, $\mathbf{b}_i$ $(i = 1, 2)$	Weight matrices and bias vectors of the NN operator
$\widetilde{\mathbf{W}}_i$, $\widetilde{\mathbf{b}}_1$ $(i = 1, 2)$	Weight matrices and bias vector of RM-IdEG
$\mathbf{W}$, $\mathbf{h}$, $\mathbf{b}$	Parameters of attentional embedding aggregator
$\mathscr{T}$, $\mathscr{T}^c$	Item sets of $\mathscr{D}^w \cup \mathscr{D}^c$ and $\mathscr{D}^c$

Note that $\mathbf{e}_t^*$ and $\mathbf{e}_t'$ do not have to be explicitly computed, and we are only interested in their gradients on $\text{IdEG}_{\text{meta}}$. Finally, we sum the two losses to get our meta-loss function:

$$l_{\text{meta}} = l_c + \eta l_w. \tag{5}$$

Here, η is the tradeoff parameter. In other words, minimizing l_{meta} simultaneously achieves two goals: (1) the error in predictions for the new items should be small; (2) after a small amount of labeled data is collected, a few gradient descent updates should lead to good generalization ability.

3. Proposed Model

3.1. Deep Interest-Shifting Network. In this section, we present DisNet, a learning framework for recommending items in a fresh-item recommendation page, which usually contains rich scenario-specific contexts. The overall network architecture is shown in Figure 2.

We note that the latent vector of a user actually reflects his or her interest in a latent space, while the scenario-specific contexts reflect the interest shifting in the user's general interests [34, 35]. For example, there is a boy who is interested in sports, games, and electronic products. Once he clicks a fresh item iPhone-11, he may pay more attention to electronic products with advanced technology, and we can recommend him newly released smartphones, laptops, and so on. We assume that such interest shifting will not change its latent semantics. In other words, the shifted representations can directly be fed into the decision-making network g. By this assumption, we can decouple the general interest of users from the scenario-specific contexts. Denoting the scenario-specific context by $\mathbf{c}_s$, we propose an interest-shifting operator (ISO) to obtain a shifted user representation:

$$\begin{aligned} \mathbf{q}_u^s &= \text{ISO}(\mathbf{q}_u, \mathbf{z}_s), \\ \mathbf{z}_s &= h(\mathbf{c}_s), \end{aligned} \tag{6}$$

where $\mathbf{q}_u^s$ and $\mathbf{q}_u$ have the same dimension m. h maps the contexts to a latent space to extract their critical information.

It is noteworthy that there is a huge amount of auxiliary data, from which we can model the general interest of the users. Thus, we can pretrain the item/user representation networks as well as the decision-making network using these data. We denote the pretrained networks by $\overline{f}_t$, $\overline{f}_u$, and $\overline{g}$. Then, the context information can be incorporated to shift the latent user vector to a scenario-specific one but in the same interest space. Formally, DisNet makes the decision by

$$\widehat{y}^{\text{DN}} = \overline{g}(\mathbf{q}_u^s, \mathbf{p}_t, \mathbf{c}) = \overline{g}\left(\text{ISO}\left(\overline{f}_u(\mathbf{v}_u), \mathbf{z}_s\right), \overline{f}_t(\mathbf{e}_t, \mathbf{v}_t), \mathbf{c}\right). \tag{7}$$

Such a model not only transfers knowledge from a general interest domain that has rich data samples but also reduces the size of the trainable parameters to the $\text{ISO}(\cdot)$ and h functions only. Obviously, the context-aware and data-deficiency problems can be addressed simultaneously.

Note that c is some contexts shared by the two domains. However, it is possible that auxiliary data have their own context as well. We ignore such contextual information and preserve the common parts only because we are modeling the general interest of the users. In practice, we also enable the decision-making network $\overline{g}$ and the embedding layer to be fine-tuned.

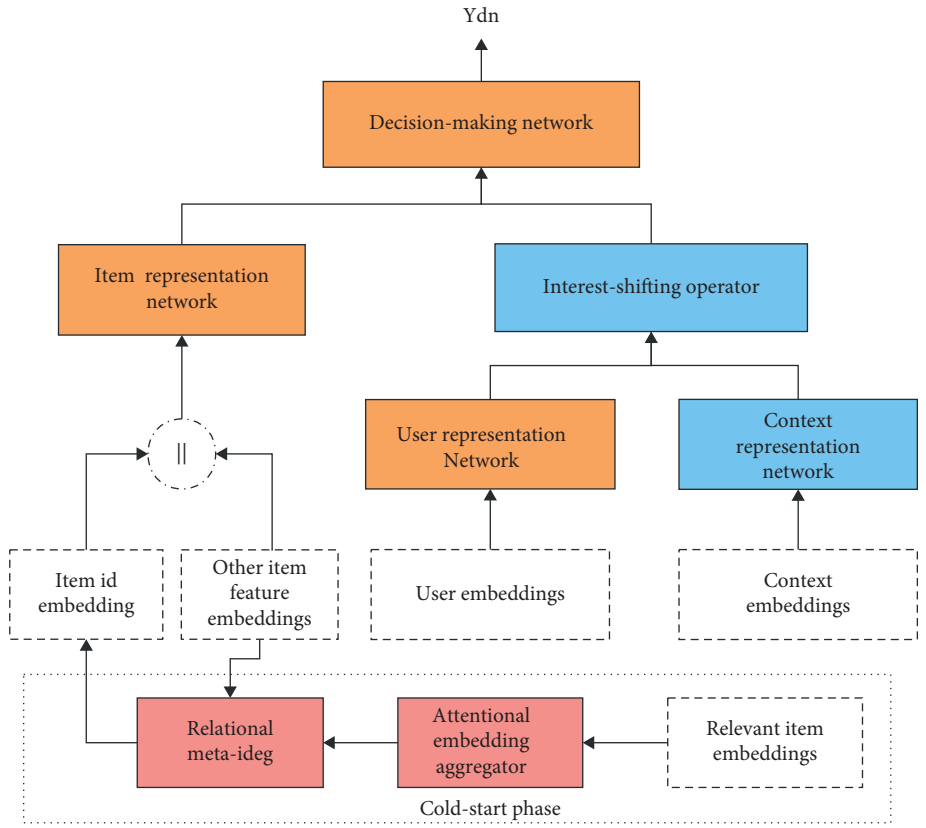


Figure 2: The model architecture of DisNet and RM-IdEG. The orange parts are pretrained using the auxiliary dataset. The blue ones are trained using the warm-start dataset. The pink parts are trained using the cold-start dataset.

3.1.1. Interest-Shifting Operators. The above discussion provides the overall network architecture. Now, we can perform any reasonable shifting operation to learn the context-specific representation of the user. In this work, we introduce three interest-shifting operators, all of which relate to very interesting interpretations.

Add Operator. Motivated by the huge success of the representation learning and knowledge graph, we adopt a similar strategy as TransR [36]. Specifically, it embeds each entity and relation by optimizing the translation principle $\mathbf{e}_a + \mathbf{e}_r \approx \mathbf{e}_b$ if a triplet (a, r, b) exists in the graph. Recall the example of interest shifting, i.e., when a boy clicks an item iPhone-11, the interest representation of this boy goes to the interest of a boy who has a preference for electronic products with advanced technology. If we regard the contextual information as a relation, we obtain our first operator, which adds up the latent user vector and contextual vector:

$$\mathrm{ISO}(\mathbf{q}_u, \mathbf{z}_s) = \mathbf{q}_u + \mathbf{z}_s. \tag{8}$$

This implies that $\mathbf{q}_u$ and $\mathbf{z}_s$ have the same dimension. That is, the projection function h directly learns the discrepancy between the original interest and the shifted interest, which is similar to the relation embedding in the knowledge graph.

COT Operator. Before introducing the second operator, we review a popular technique in the context-aware recommendation, namely, the contextual operation tensor (COT) [37]. By estimating a contextual operation matrix, COT maps the original user/item latent vectors to their context-specific ones. We notice that COT has three main limitations: (1) it assumes the context space is fixed and the contextual operation matrix relates to different context values; (2) it jointly learns the original latent vectors as well as the contextual operation matrix; and (3) it uses linear mapping, i.e., a 3D tensor, to obtain the contextual operation matrix, which leads to degenerated performance. Obviously, COT cannot be applied to our problem directly because the data-deficiency problem prevents the joint learning procedure, and cross-domain data have different contexts.

Fortunately, in DisNet, we have decoupled the user's general interest from the scenario-specific interest. Therefore, we can estimate the scenario-specific context operation matrix using the h function:

$$\begin{aligned} \mathrm{ISO}(\mathbf{q}_u, \mathbf{Z}_s) &= \mathbf{Z}_s^{\top} \mathbf{q}_u, \\ \mathbf{Z}_s &= h(\mathbf{c}_s). \end{aligned} \tag{9}$$

Here, h outputs a $d \times d$ matrix instead of a single vector. In other words, while COT focuses on different context

values, our model considers how external contexts affect the user's interest.

Neural Network-Based Operator. Yet, we have only considered linear shifting, while in reality, the transformation may be nonlinear. To bridge this gap, we propose a neural network-based operator:

$$\mathrm{ISO}\left(\mathbf{q}_u, \mathbf{z}_s\right) = \mathbf{W}_2^\top \sigma\left(\mathbf{W}_1^\top \left[\mathbf{q}_u \| \mathbf{z}_s\right] + \mathbf{b}_1\right) + \mathbf{b}_2, \tag{10}$$

where $\mathbf{W}_i$, $\mathbf{b}_i$ $(i \in \{1, 2\})$ refer to the weight matrices and bias vectors. σ is the activation function. $[\cdot\|\cdot]$ denotes the concatenation of two vectors. It is worth pointing out that any network architectures can be used, and this paper considers a simple multilayer perceptron.

While the add operator regards the contexts as bias and the COT operator considers the cross-influences between the user interest and contexts, the NN-based operator achieves these two goals simultaneously.

3.2. Relational Meta-Id-Embedding Generator. This section concentrates on babysitting fresh items in the cold-start phase, where they suffer from a severe cold-start problem. It is worth noting that DisNet can reuse the embedding layer after pretraining. Then, all the attributes except item id obtain great embeddings. Hence, the only thing that matters is the item id embedding initialization. Following [27], this work learns an IdEG in a learning-to-learn manner. Nevertheless, we notice that the vanilla meta-IdEG feeds item features into a simple neural network to generate embeddings. Obviously, meta-IdEG neglects the fact that id embedding reflects the community structural information between items, exploiting which has been proved beneficial for alleviating the cold-start problem [26].

To remedy this problem, a novel relational meta-Id embedding generator (RM-IdEG) is proposed, whereas it trains the item id embedding in a learning-to-learn manner and integrates relational information for better embedding initialization, which further improves the performance of DisNet on new items. Specifically, we collect a set of warm-start items that are significantly relevant to the cold-start item t. Many influential relations can be considered, such as items from the same seller and the same brand. For instance, a newly released Nike T-shirt may have similar selling behaviors as other items in Nike shops. Then, we construct an id embedding set $\mathcal{J}_t = \{\mathbf{e}_t^1, \ldots, \mathbf{e}_t^k\}$. Here, $\mathbf{e}_t^i$ $(i = 1, 2, \ldots, k)$ denote the id embeddings of top k relevant items. Then, we output the new embedding via an attentional embedding aggregator:

$$\begin{aligned} \mathbf{e}_t^* &= \frac{1}{C}\sum_{i=1}^{k} a_t^i \mathbf{e}_t^i, \\ C &= \sum_{i=1}^{k} a_t^i. \end{aligned} \tag{11}$$

Here, C is used for normalization. The attention score a_t^i is given by a global attention network:

$$a_t^i = \exp\left(\mathbf{h}^T \sigma\left(\mathbf{W}^\top \mathbf{e}_t^i + \mathbf{b}\right)\right), \tag{12}$$

where $\mathbf{h}$, $\mathbf{W}$, and $\mathbf{b}$ are shared attention parameters. Then, we feed the learned attentional id embedding and item features into a neural network to obtain the final embedding:

$$\mathbf{e}_t = \underset{\mathrm{rm}}{\mathrm{IdEG}}\left(\mathbf{v}_t, \mathcal{J}_t\right) = \tanh\left(\widetilde{\mathbf{W}}_2^\top \sigma\left(\widetilde{\mathbf{W}}_1^\top \left[\mathbf{e}_t^* \| \mathbf{v}_t\right] + \widetilde{\mathbf{b}}_1\right)\right), \tag{13}$$

where $\widetilde{\mathbf{W}}_i$ $(i \in \{1, 2\})$ are weight matrices and $\widetilde{\mathbf{b}}_1$ is the bias vector. To obtain numerically stable outputs, we follow some tricks in [27]: (1) the bias of the last layer is removed; (2) tanh activation is applied in the final layer.

Remark 1. The proposed model fully addresses the cold-start problem from two aspects: (1) through a learning-to-learn training procedure, our model achieves better generalization ability with few training data; (2) by considering influentially relevant items, RM-IdEG automatically encodes community structural information into the embedding initialization, and the predictive accuracy is further improved.

3.3. Training. Now, we describe the training procedure of our model. Note that the training fresh item set $\mathcal{T}$ does not contain those newborn items. Consequently, we choose an item subset $\mathcal{T}^c$ from $\mathcal{T}$ to simulate the cold-start setting. For each item in $\mathcal{T}^c$, which corresponds to a task, we preserve m examples for both the support set and the query set (a total of $2m$ examples). The remaining examples of these items are dropped since they should not appear before we train the RM-IdEG. To avoid the performance of the base model being decreased, we limit each item in $\mathcal{T}^c$ to having less than or equal to M examples $(M > 2m)$ and obviously, greater or equal to $2m$ examples. We denote the constructed cold-start dataset by $\mathcal{D}^c$. The data examples of the remaining items $\mathcal{T} - \mathcal{T}^c$ constitute the warm-start dataset $\mathcal{D}^w$. Remark that the items in $\mathcal{T}$ are all warm-start items since they have at least one data example. We call $\mathcal{D}^c$ cold-start because they are used to train RM-IdEG, which is designed for totally cold-start items. Also, $\mathcal{D}^w$ is called warm-start since it is used to train DisNet, which does not consider the cold-start problem.

In summary, we have three datasets: (1) an auxiliary dataset $\mathcal{D}^a$, having no scenario-specific contexts, collected from other domains; (2) a warm-start dataset $\mathcal{D}^w$ that has rich contextual information; and (3) a cold-start dataset $\mathcal{D}^c$ that contains few-shot examples. Accordingly, the whole model is trained in three stages, and we put the details in Algorithm 1.

4. Experiments

To justify the effectiveness of DisNet and RM-IdEG, we conduct comprehensive experiments to answer the following questions:

RQ1: can DisNet outperform state-of-the-art methods?

RQ2: can RM-IdEG outperform state-of-the-art IdEGs?

RQ3: is our model sensitive to the parameters?

Input: $\mathcal{D}^a$: auxiliary dataset
Input: $\mathcal{D}^w$: warm-start dataset
Input: $\mathcal{D}^c$: cold-start dataset
Input: $(\tilde{t}, \tilde{u}, \mathbf{c}, \mathbf{c}_s)$: a testing example
Output: $\tilde{y}^{\mathrm{DN}}$: the predicted label of $(\tilde{t}, \tilde{u}, \mathbf{c}, \mathbf{c}_s)$

repeat
  /* the first stage, pretrain the model using auxiliary data */
  Randomly sample a batch of data from $\mathcal{D}^a$
  Calculate the predicted label $\hat{y}$ by equation (1)
  Update g, f_u, f_t by gradient descent
until *Converge*
Fix g, f_u, f_t to $\overline{g}$, $\overline{f}_u$, $\overline{f}_t$
repeat
  /* the second stage, train DisNet using warm-start data */
  Randomly sample a batch of data from $\mathcal{D}^w$
  Calculate $\mathbf{q}_u$, $\mathbf{p}_t$ using $\overline{f}_u$, $\overline{f}_t$
  Compute the shifted interest vector $\mathbf{q}_u^s$ by equation (6)
  Calculate the predicted label $\hat{y}^{\mathrm{DN}}$ using $\overline{g}$ by equation (7)
  Update h, $\mathrm{ISO}(\cdot)$ by gradient descent
until *Converge*
Fix all the trainable parameters except the item id embeddings
repeat
  /* the third stage, train RM-IdEG using cold-start data */
  Randomly sample an item t_i and get its support/query sets $(\mathcal{D}_i^s, \mathcal{D}_i^q)$ from $\mathcal{D}^c$
  Aggregate embeddings of relational items of t_i by equation (11)
  Generate an id embedding $\mathbf{e}_t^*$ for t_i using RM-IdEG
  Compute the cold-start loss on $\mathcal{D}_i^s$ by equation (2)
  Update the id embedding of t_i to $\mathbf{e}_t'$ by equation (3)
  Compute the warmed loss on $\mathcal{D}_i^q$ by equation (4)
  Update RM-IdEG by gradient descent
until *Converge*
if $\tilde{t}$ is a cold-start item **then**
  Generate an id embedding $\tilde{\mathbf{e}}_t$ for $\tilde{t}$ using RM-IdEG
else
  Get the id embedding $\tilde{\mathbf{e}}_t$ of $\tilde{t}$ from the embedding layer
end if
Return a label $\tilde{y}^{\mathrm{DN}}$ for $(\tilde{t}, \tilde{u}, \mathbf{c}, \mathbf{c}_s)$ by equation (7) using DisNet

Algorithm 1: The training and testing procedure.

4.1. Dataset

4.1.1. Dataset Description. We evaluate our methods on two synthetic datasets and a real-world dataset:

MovieLens (https://grouplens.org/datasets/movielens/) [38]: it consists of 1.0 million movie-ranking instances across about 6,000 users and 4,000 movies. The features of movies include movie id, title, year of release, and genres. Titles and genres are lists of tokens. The features of users include user id, age, gender, occupation, and zipcode. To simulate our fresh item setting, we choose gender, occupation, and zipcode as scenario-specific context features. We also convert the rating scores to binary values. The ratings smaller than 4 are turned into 0, and the others are turned into 0.

Book-Crossing (http://www2.informatik.uni-freiburg.de/cziegler/BX/) [39]: it is collected by Cai-Nicolas Ziegler in a one-month crawl from the Book-Crossing (http://www.bookcrossing.com/) community. It contains 0.27 million users, providing 1.15 million ratings about 0.28 million books. The features of books include ISBN number (book id), book title, year of publication, and publisher. The features of users include age and location. Similar to MovieLens, we select location as a scenario-specific context feature. The ratings are converted to 1 if they are at least 4 and 0, otherwise.

Taobao-Fresh: it collects 203.1 million user-item click interactions produced by the main entrance page of Taobao's app as auxiliary data and 4.4 million user-item click interactions produced by the *New Tendency* page as fresh-item recommendation data. A total of 4.8 million users and 1.6 million items are considered, with 71 user features, 17 item features, and 17 contextual features (auxiliary data have no contexts).

4.1.2. Data Splitting. For MovieLens and Book-Crossing, we first group the items by their ids. We put those items with the number of examples less than $M + 1$ and larger than $2m - 1$ in $\mathcal{T}^c$. Then, we construct a cold-start dataset $\mathcal{D}^c$ by

preserving $2m$ examples for each item. From the examples of the remaining items $\mathcal{T} - \mathcal{T}^c$, we randomly choose 80% as auxiliary data $\mathcal{D}^a$ and 20% as the warm-start dataset $\mathcal{D}^w$. We set $m = 20$ and $M = 100$ for MovieLens. For Book-Crossing, we notice that a total of 48,434 books are rated by exactly 2 users. Hence, we set $m = 1$ and $M = 2$. It enables us to study an extreme experimental setting, i.e., each cold-start item is one-shot.

For Taobao-Fresh, the auxiliary data $\mathcal{D}^a$ have been collected. We then split the fresh-item recommendation data into two parts. The first one is a cold-start dataset $\mathcal{D}^c$ where items have greater than or equal to 10 interactions and less than or equal to 20 interactions. Similarly, each item in $\mathcal{D}^c$ has a support set and a query set, each of which has 5 examples. The examples of the remaining items are collected as the warm-start dataset $\mathcal{D}^w$. The statistics of these datasets can be found in Table 2.

4.1.3. Data Generation. To answer RQ1, for each dataset, we run DisNet on three types of data:

Auxiliary-only data: they contain the auxiliary data and context-free warm-start data, i.e., the context features of the warm-start data are removed.

Context-only: it is exactly warm-start data. In other words, DisNet is run without pretraining.

Full data: they comprise auxiliary data and warm-start data and are the main setting of this paper.

Note that the three types of data are used to test the effectiveness of DisNet, while cold-start data are used to evaluate the superiority of RM-IdEG.

For performance evaluation, we randomly divide the warm-start and cold-start data into 80% training and 20% testing. We run the experiments for five times, and the mean AUC performance on the testing set is reported.

4.2. Baselines. We evaluate the proposed model in two stages. In the first stage, we compare DisNet with three context-aware recommendation models:

DeepFM [11]: it feeds embeddings to a factorization machine model as well as a multilayer perceptron and then aggregates their outputs and gets the final prediction.

PNN [13]: the dense embeddings are fed into a dense layer and a product layer. Then, it concatenates their outputs together and uses a two-layer neural network to get the prediction.

CFM [15]: CFM is a recent state-of-the-art CARS method that explicitly learns second-order feature interactions. It calculates the pairwise outer product of dense embeddings and stacks them to obtain an interaction cube. Then, it applies the convolution pooling technique to get the final prediction.

The dimension of embedding vectors of each input field is fixed to 128, and the activation function is chosen as ReLU for all the models. As suggested in [11], we use three dense hidden layers as the deep component for both DeepFM and PNN. For DisNet, the size of the user/item latent representation is set as 64. We use two fully connected layers with a hidden dimension of 64 for user/item representation networks as well as the decision-making network. We do not activate the outputs of user/item representation networks. The context network of the NN/add ISO and the shifting network of the NN ISO also comprise two fully connected layers with hidden size 64 and without activation in the final layer. For the COT ISO, we linearly learn a contextual operation matrix of size 64×64 from the contexts. Finally, the learning rate and l_2-regularization parameters are fine-tuned by five-fold cross-validation.

Then, we evaluate the RM-IdEG with two baselines:

Rand-IdEG: the random initialization of id embeddings is one of the most commonly used strategies in recommender systems.

Meta-IdEG [27]: the state-of-the-art solution to the cold-start problem. It firstly feeds the item features into a simple neural network to generate embeddings and then trains them in a learning-to-learn manner.

For Rand-IdEG, we initialize the id embeddings with random values from a standard Gaussian distribution with standard deviation 0.01. For meta-IdEG, we use the neural network architecture as suggested in [27]. For RM-IdEG, we use a two-layer neural network with a hidden size of 128 as the IdEG network. According to Pan et al. [27], the tradeoff parameter η is robust. Hence, we follow their experimental setting and set η as 0.1 for meta-IdEG and RM-IdEG. We also follow their two suggestions that use tanh as activation and remove the bias of the output layer. For a target item in the synthetic dataset, we choose k-nearest neighbors from the previous training dataset, i.e., $\mathcal{D}^w \cup \mathcal{D}^a$, using hamming distance as the relevant items, where k is chosen by five-fold cross-validation. For Taobao-Fresh, we randomly select 10 items having the same seller and 10 items having the same brand as the relevant items. We choose DisNet-NN as the base model, which has been pretrained by $\mathcal{D}^a$ and $\mathcal{D}^w$.

4.3. Empirical Results

4.3.1. Performance Comparison of Context-Aware Models (RQ1). Tables 3 and 4 report the testing AUC comparison of three context-aware models on two synthetic datasets and the Taobao-Fresh dataset. We have the following findings:

All the methods obtain the best performance on the full data. For example, on Taobao-Fresh, DisNet-NN improves the AUC scores on auxiliary-only and context-only data by 1.00% and 1.69%, respectively. This finding verifies the importance of utilizing auxiliary data and contexts to alleviate the data-deficiency problem.

On the Taobao-Fresh dataset, all the methods achieve significantly greater improvement on the context-only data than the auxiliary-only data. It demonstrates that, in the fresh-item recommendation task, the context information highly reflects the user interest.

Table 2: The statistics of the datasets.

Datasets	# users	# items	# interactions (M)†	# auxiliary (M)	# warm-start (M)	# cold-start‡
Book-Crossing	0.28 M	0.27 M	1.15	0.75	0.19	48,434
MovieLens	6,040	3,706	1.00	0.76	0.19	30,645
Taobao-Fresh	4.78 M	1.61 M	207.54	203.13	4.09	0.31 M

†The first three columns list the number of users, items, and data examples of the whole dataset. ‡The second three columns report the number of data examples of the preprocessed datasets after splitting.

Table 3: Testing AUC comparison of context-aware models on synthetic datasets.

Methods	Dataset: Book-Crossing			Dataset: MovieLens		
	Auxiliary†	Context	Full data	Auxiliary†	Context	Full data
DeepFM	0.7836•	0.7741•	0.7840•	0.7284•	0.7586•	0.7654•
PNN	0.7854•	0.7725•	0.7857•	0.7290•	0.7584•	0.7649•
CFM	0.7730•	0.7726•	0.7707•	0.7289•	0.7571•	0.7633•
DisNet-Add	–	0.7745°	0.7862•	–	0.7576°	0.7660•
DisNet-COT	–	0.7733•	0.7864•	–	0.7559•	0.7664°
DisNet-NN	0.7858	0.7748	0.7878	0.7287	0.7577	0.7666

•/° indicates whether the DisNet-NN variant is significantly superior to the coupling algorithm or not on each dataset (pairwise t-test at the 0.05 significance level).†On the auxiliary-only data, the network architecture of DisNet is fixed, and we only report the performance once.

On auxiliary-only data, all the models are competitive with each other. However, on full data, the performance of baselines shows no significant improvement after the context features being involved. The reason is these baselines deeply couple the context in the model, and thus, the knowledge of the auxiliary domain cannot be fully utilized. Take DeepFM as an example; since $\mathscr{D}^a$ and $\mathscr{D}^w$ have different input formats, the deep component cannot be reused. Though we can reuse the embedding layer, its predictive performance is limited.

DisNet models with full data significantly outperform all the baselines as well as their auxiliary-only and context-only counterparts. The interest-shifting operator enables us to completely exploit both context and cross-domain information.

Different interest-shifting operators show competitive performance with each other. Moreover, the NN-based operator obtains the best performance because it enables the user interest to be shifted nonlinearly.

Interestingly, DisNet-COT always underperforms DisNet-Add on the context-only dataset but is better than DisNet-Add on the full dataset. We suppose the reason is the COT operator tends to overfit on context-only data since it contains more parameters. With the help of auxiliary data, this problem is alleviated.

4.3.2. Performance Comparison of Different IdEGs (RQ2). Tables 5 and 6 list the cold-start and warmed-up performance of DisNet with different id embedding generators. Once the IdEG produces the id embeddings, the cold-start performance is directly evaluated on a meta-testing query set, where all items are cold-start ones. Then, we perform one step of gradient descent to update the id embeddings using a meta-testing support set that contains the same items as the query set. Finally, the warmed-up performance is evaluated again on the query set.

Table 4: Testing AUC comparison of context-aware models on the Taobao-Fresh dataset.

Methods	Auxiliary†	Context	Full data
DeepFM	0.7367•	0.7449•	0.7362•
PNN	0.7413•	0.7439•	0.7417•
CFM	0.7377•	0.7441•	0.7442•
DisNet-Add	–	0.7480°	0.7528•
DisNet-COT	–	0.7467•	0.7533°
DisNet-NN	0.7409	0.7483	0.7534

•/° indicates whether the RM-IdEG variant is significantly superior to the coupling algorithm or not (pairwise t-test at the 0.05 significance level).†On the auxiliary-only data, the network architecture of DisNet is fixed, and we only report the performance once.

From the results, we conclude that

Meta-IdEG and RM-IdEG outperform Rand-IdEG on both cold-start and warmed-up phases because the learning-to-learn training procedure guarantees them to quickly achieve good generalization ability on unseen data.

RM-IdEG achieves the best performance on all the datasets. In particular, even with one-shot training, RM-IdEG still outperforms on the Book-Crossing dataset. By integrating information of significantly relevant items, RM-IdEG inherently models the community structural information when initializing id embeddings.

4.3.3. Parameter Sensitivities (RQ3). The main parameters are the tradeoff parameter of the meta-loss η and the number of relevant items k. The robustness of η has been studied in [27]. Thus, we investigate the sensitivity of k and the results on Book-Crossing and MovieLens datasets which are shown in Figure 3. We can see that when k is small, the performance is close to Meta-IdEG because few relational information is learned. The best result is obtained when $k = 6$, and then the

Table 5: Testing AUC comparison of different IdEGs on synthetic datasets.

IdEG types	Dataset: Book-Crossing		Dataset: MovieLens	
	Cold-start	Warmed-up	Cold-start	Warmed-up
Rand-IdEG	0.7940 (+0.00%)•	0.7943 (+0.00%)•	0.7065 (+0.00%)•	0.7377 (+0.00%)•
Meta-IdEG	0.7945 (+0.06%)•	0.7948 (+0.06%)•	0.7132 (+0.95%)•	0.7680 (+4.11%)•
RM-IdEG	0.7951 (+0.14%)	0.7955 (+0.15%)	0.7174 (+1.54%)	0.7735 (+4.85%)

•/° indicates whether the DisNet-NN variant is significantly superior to the coupling algorithm or not on each dataset (pairwise t-test at the 0.05 significance level).

Table 6: Testing AUC comparison of different IdEGs on the Taobao-Fresh dataset. The best ones are shown in bold.

IdEG types	Cold-start	Warmed-up
Rand-IdEG	0.5792 (+0.00%)•	0.6042 (+0.00%)•
Meta-IdEG	0.6133 (+5.89%)•	0.6361 (+5.28%)•
RM-IdEG	**0.6160 (+6.35%)**	**0.6382 (+5.63%)**

•/° indicates whether RM-IdEG is significantly superior to the coupling algorithm or not (pairwise t-test at the 0.05 significance level).

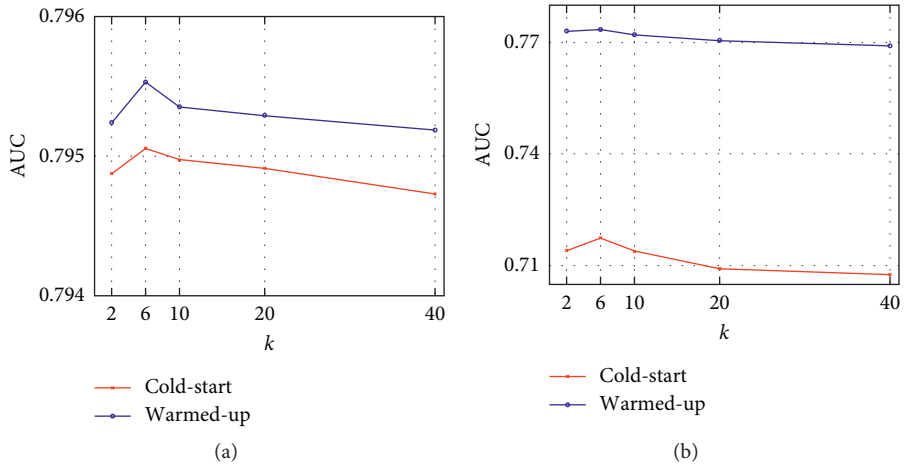


Figure 3: Parameter sensitivity of k on two datasets. (a) Book-Crossing. (b) MovieLens.

performance drops. The reason is that, as k becomes larger, the relations become weaker, but the model complexity increases.

5. Related Work

5.1. Context-Aware Recommendation. Context-aware recommender systems (CARSs) have attracted considerable attention in past years [7]. Early work in CARS can be divided into two categories: (1) prefiltering methods [40], where context guides the selection of training data; (2) postfiltering methods [41], where context drives recommendation results' selection. The main limitation of these methods is that they require the supervision and fine-tuning in all steps of recommendation [42]. To address this problem, contextual modeling approaches capture the contextual information directly in model construction. Some works are based on matrix factorization [8], such as CAMF [28] and CSLIM [9]. Another group of studies exploits tensor factorization techniques for modeling user-item-context relations [43, 44]. Recently, factorization machines [42, 45, 46] and deep learning [47, 48] based on CARS become increasingly popular, which directly model nonlinear interactions between features. Some studies also use representation learning techniques, e.g., CARS^2 [49] and COT [37], which provide not only a latent vector but also context-aware representations. In summary, all the above methods assume the data are sufficient for training, while severe data-deficiency problem occurs in many fresh-item recommendation pages.

5.2. Cross-Domain Recommendation. As we have discussed, data deficiency is one of the most challenging problems for recommender systems, and it is much more significant in many fresh-item recommendation scenarios. One promising solution to this problem is cross-domain recommender systems (CDRSs) [50]. Existing CDRSs can be categorized into symmetric and asymmetric ones. Symmetric models [16, 18, 51, 52] collect sparse data from multiple domains and anticipate that these domains can complement each other. In our task, symmetric strategy is incompatible because the two domains have heterogeneous data format and imbalance data size. Thus, we consider asymmetric models [19, 20, 21],

which aim to leverage data in an auxiliary domain to alleviate data deficiency of the target domain. In this way, knowledge learned from the auxiliary domain is directly transferred to the target domain, acting as priors or regularization. Nevertheless, many asymmetric CDRSs adopt shallow methods and have the difficulty in learning complex user-item interaction relationship [18, 26]. Moreover, scenario-specific contextual information of the target domain has been seldom considered.

5.3. Cold-Start Recommendation. When recommending cold-start fresh items, a severe cold-start problem occurs. To handle this problem, it is common to collect information for the cold item or user, e.g., item attributes [22, 23] and user attributes [24, 25]. A recent work HERS [26] also utilizes relational data to boost performance, such as social information of users. In [16], the authors explored a symmetric cross-domain recommender system, where shared knowledge can help alleviate the cold-start problem.

Recently, a series of works [27, 53, 54] also adopt meta-learning technique [55] which enables the recommender system to achieve good generalization ability after few-shot training. From the cold-start user perspective, MeLU [53] learns a meta-id embedding for the cold-start users and then predicts the user preference on the items by the norm of gradients. From the cold-start item perspective, Pan et al. [27] proposed the meta-Id embedding generator (meta-IdEG), which also takes id embedding initialization into account. However, since meta-IdEG only uses item features to generate id embedding, it ignores the community structural information concealed in id embedding, which leads to a suboptimal solution.

6. Discussion and Conclusion

6.1. Further Discussion. In this section, we discuss the significance of this work.

6.1.1. Importance of the Application. The fresh-item recommendation task reveals a new perspective of personalized recommendation, i.e., the impact of items' life period. Some people may prefer products which stand the test of time, while some others may be interested in newly released products. The *New Tendency* page enables the latter ones' recommendation to be fully personalized. From another point of view, these fresh items also obtain more opportunities to be exposed. Hence, high-quality and novel products can quickly become popular. We also address the main difficulties of this learning task, i.e., data deficiency and cold-start.

6.1.2. Importance of the Techniques. Surprisingly, though the two techniques DisNet and RM-IdEG are proposed to handle the fresh-item recommendation task, we find that both methods have a wide range of applications.

As aforementioned, the DisNet is designed for fresh-item recommendation pages. Actually, such pages are quite common in existing E-commerce platforms. For example, after a bill being paid, the E-commerce platform will recommend other related items to the customers. It is a classical fresh-item recommendation scenario. Obviously, a fresh-item recommendation page usually contains rich contextual information. The contexts reflect that the user interest shifts from a general one to a scenario-specific one. However, with fewer page views, such pages usually face severe data-deficiency problems. And this work can address this issue by giving a novel learning framework, which simultaneously transfers knowledge from an auxiliary domain as well as fully utilizes the context information.

RM-IdEG can also be applied to many real-world applications. In [27], the authors proposed to learn meta-id-embeddings for cold-start advertisements. And we can also collect relevant advertisements by its company, topic, and so on. As a result, the model can generate better id embeddings. Furthermore, other relational data can also be considered. For instance, if we consider the user cold-start problem [53], we may explore the social networks of a new user so that RM-IdEG is able to initialize a fast-adapting and relation-aware id embedding.

6.2. Conclusion. In this work, we address two difficulties of the fresh-item recommendation task. First, we propose a deep interest-shifting network to deal with the data-deficiency problem of fresh item recommendation. Specifically, users' general interests are learned from a huge number of an auxiliary dataset. Then, our model shifts the user interest to a scenario-specific one using context features. Second, we propose a relational meta-Id-embedding generator (RM-IdEG) to alleviate the cold-start problem. RM-IdEG is trained in a learning-to-learn manner with relational information being integrated. Hence, community structural information can be inherently embedded in the id embeddings of newborn items. Extensive experiments on two synthetic datasets and a real-world dataset clearly identify the effectiveness of our approaches, which have been already deployed on a large-scale online fresh-item recommendation application.

Data Availability

Previous reported data were used to support this study, and these prior studies (and datasets) are cited at relevant places within the text as references [38, 39].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 61972336).

References

[1] J. Song, Li Zhao, Z. Hu et al., "Poisonrec: an adaptive data poisoning framework for attacking black-box recommender systems," in *Proceedings of the ICDE 2020*, pp. 157–168, Dallas, TX, USA, 2020.

[2] P. Wang, Li Zhao, X. Pan, D. Ding, X. Chen, and Y. Hou, "Density matrix based preference evolution networks for e-commerce recommendation," in *DASFAA*, pp. 366–383, Springer, Berlin, Germany, 2019.

[3] P. Wang, Li Zhao, Y. Zhang, Y. Hou, and L. Ge, "QPIN: a quantum-inspired preference interactive network for E-commerce recommendation," in *Proceedings of the CIKM*, pp. 2329–2332, Beijing, China, 2019.

[4] J. Chen, Z. Wang, T. Zhu, and F. E. Rosas, "Recommendation algorithm in double-layer network based on vector dynamic evolution clustering and attention mechanism," *Complexity*, vol. 2020, no. 2020, 19 pages, Article ID 5206087.

[5] Z. Li, F. Xiong, X. Wang, H. Chen, and Xi Xiong, "Topological influence-aware recommendation on social networks," *Complexity*, vol. 2019, Article ID 6325654, 12 pages, 2019.

[6] Li Zhao, L. Zhang, C. Lei, X. Chen, J. Gao, and J. Gao, "Attention with long-term interval-based deep sequential learning for recommendation," *Complexity*, vol. 2020, Article ID 6136095, 13 pages, 2020.

[7] G. Adomavicius, B. Mobasher, F. Ricci, and T. Alexander, *Context-Aware Recommender Systems*, AI Magazine, Palo Alto, CA, USA, 2011.

[8] M. H. Abdi, O. George, and R. W. Mwangi, "Matrix factorization techniques for context-aware collaborative filtering recommender systems: a survey." *Computer and Information Science*, vol. 11, no. 2, 2018.

[9] Y. Zheng, B. Mobasher, D Robin, and Burke, "Deviation-based contextual SLIM recommenders," in *Proceedings of the CIKM*, pp. 271–280, Shanghai, China, November 2014.

[10] S. Wang, J. Zhang, and X.D. Zhang, "A time-aware CNN-based personalized recommender system," *Complexity*, vol. 2019, Article ID 9476981, 11 pages, 2019.

[11] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "DeepFM: a factorization-machine based neural network for CTR prediction," in *Proceedings of the IJCAI 2017*, pp. 1725–1731, Melbourne, Australia, 2017.

[12] L. Guo, H. Jiang, X. Liu, and C. Xing, "Network embedding-aware point-of-interest recommendation in location-based social networks," *Complexity*, vol. 2019, Article ID 3574194, 18 pages, 2019.

[13] Y. Qu, B. Fang, W. Zhang et al., "Product-based neural networks for user response prediction over multi-field categorical data," *ACM Transactions on Information Systems*, vol. 37, no. 1, 2019.

[14] O. Tal and Y. Liu, "A joint deep recommendation framework for location-based social networks," *Complexity*, vol. 2019, Article ID 2926749, 11 pages, 2019.

[15] X. Xin, Bo Chen, X. He, D. Wang, Y. Ding, and J. Jose, "CFM: convolutional factorization machines for context-aware recommendation," in *Proceeding of the IJCAI*, pp. 3926–3932, Macao, China, 2019.

[16] Da Cao, X. He, L. Nie et al., "Cross-platform app recommendation by jointly modeling ratings and texts," *ACM Transactions on Information Systems*, vol. 35, no. 4, 2017.

[17] C. Gao, X. Chen, F. Feng et al., "Cross-domain recommendation without sharing user-relevant data," 2019.

[18] G. Hu, Yu Zhang, and Q. Yang, "CoNet: collaborative cross networks for cross-domain recommendation," in *Proceedings of the CIKM 2018*, pp. 667–676, Turin, Italy, 2018.

[19] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu, "Personalized recommendation via cross-domain triadic factorization," 2013.

[20] B. Li, Q. Yang, and X. Xue, "Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction," in *Proceedings of the IJCAI*, pp. 2052–2057, Pasadena, CA, USA, 2009.

[21] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction," in *Proceedings of the AAAI*, Atlanta, GA, USA, 2010.

[22] S. Martin and M. Amin, "Item cold-start recommendations: learning local collective embeddings," in *Proceedings of the RecSys*, pp. 89–96, Foster City, CA, USA, 2014.

[23] M. Volkovs, G. W. Yu, and T. Poutanen, "DropoutNet: addressing cold start in recommender systems," 2017.

[24] S. Roy and S. C. Guntuku, "Latent factor representations for cold-start video recommendation," in *Proceedings of the RecSys 2016*, pp. 99–106, Boston, MA, USA, 2016.

[25] Y. Seroussi, F. Bohnert, and I. Zukerman, "Personalised rating prediction for new users using latent factor models," 2011.

[26] L. Hu, S. Jian, L. Cao, Z. Gu, Q. Chen, and A. Amirbekyan, "HERS: modeling influential contexts with heterogeneous relations for sparse and cold-start recommendation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 3830–3837, 2019.

[27] F. Pan, S. Li, Ao Xiang, P. Tang, and Q. He, "Warm up cold-start advertisements: improving CTR predictions via learning to learn ID embeddings," , pp. 695–704, 2019.

[28] L. Baltrunas, B. Ludwig, and F. Ricci, "Matrix factorization techniques for context aware recommendation," in *Proceedings of the RecSys 2011*, pp. 301–304, Chicago, IL, USA, October 2011.

[29] H. Wang, F. Zhang, X. Xie, and M. Guo, "DKN: deep knowledge-aware network for news recommendation," 2018.

[30] Q. Zhu, X. Zhou, Z. Song, J. Tan, and L. Guo, "DAN: deep attention neural network for news recommendation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 5973–5980, 2019.

[31] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," in *Proceedings of the IJCAI*, pp. 4981–4987, Yokohama, Japan, 2020.

[32] J. Wu, S. Pan, X. Zhu, and Z. Cai, "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2014.

[33] J. Wu, X. Zhu, C. Zhang, and S. Y. Philip, "Bag constrained structure pattern mining for multi-graph classification," *Ieee Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[34] X. Chen, G. Yu, J. Wang, C. Domeniconi, Li Zhao, and X. Zhang, "ActiveHNE: active heterogeneous network embedding," in *Proceedings of the IJCAI 2019*, pp. 2123–2129, Macao, China, 2019.

[35] C. Chu, Li Zhao, B. Xin et al., "Deep graph embedding for ranking optimization in e-commerce," in *Proceedings of the CIKM*, pp. 2007–2015, Turin, Italy, 2018.

[36] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proceedings of the AAAI*, pp. 2181–2187, Austin, TX, USA, 2015.

[37] S. Wu, Q. Liu, L. Wang, and T. Tan, "Contextual operation for recommender systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 8, pp. 2000–2012, 2016.

[38] F. Maxwell Harper and J. A. Konstan, "The movielens datasets: history and context," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 4, p. 19, 2016.

[39] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," 2005.

[40] G. Adomavicius, S. Ramesh, S. Sen, and T. Alexander, "Incorporating contextual information in recommender systems using a multidimensional approach," *ACM Transactions on Information Systems*, vol. 23, p. 1, 2005.

[41] U. Panniello, T. Alexander, M. Gorgoglione, C. Palmisano, and A. Pedone, "Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems," in *Proceedings of the RecSys 2009*, pp. 265–268, Copenhagen, Denmark, 2009.

[42] W.-Y. Ma, J.-Y. Nie, R. Baeza-Yates, T.-S. Chua, and W. Bruce Croft, 2011, SIGIR: ACM.

[43] B. Hidasi and T. Domonkos, *Fast ALS-Based Tensor Factorization for Context-Aware Recommendation from Implicit Feedback*, pp. 67–82, Springer, Berlin, Germany, 2012.

[44] Y. Shi, A. Karatzoglou, L. Baltrunas, M. A. Larson, A. Hanjalic, and N. Oliver, "TFMAP: optimizing map for top-n context-aware recommendation," 2012.

[45] X. Chen, Y. Zheng, J. Wang, W. Ma, and J. Huang, "RaFM: rank-aware factorization machines," in *Proceedings of the ICML 2019*, pp. 1132–1140, Long Beach, CA, USA, 2019.

[46] B. Loni, R. Pagano, M. Larson, and A. Hanjalic, "Top-N recommendation with multi-channel positive feedback using factorization machines," *ACM Transactions on Information Systems*, vol. 37, no. 2, 2019.

[47] H.-T. Cheng, L. Koc, J. Harmsen et al., "Wide & deep learning for recommender systems," in *Proceedings of the RecSys 2016*, pp. 7–10, Boston, MA, USA, 2016.

[48] G. Zhou, X. Zhu, C. Song et al., "Deep interest network for click-through rate prediction," in *Proceedings of the KDD*, pp. 1059–1068, London, UK, 2018.

[49] Y. Shi, A. Karatzoglou, L. Baltrunas, M. A. Larson, and A. Hanjalic, "CARS2: learning context-aware representations for context-aware recommendations," in *Proceedings of the CIKM*, pp. 291–300, Shanghai, China, 2014.

[50] I. Fernández-Tobías, I. Cantador, M. Kaminskas, and F. Ricci, "Cross-domain recommender systems: a survey of the state of the art," in *Proceedings of the Spanish Conference on Information Retrieval*, Valencia, Spain, 2012.

[51] C.-Y. Li and S.-D. Lin, "Matching users and items across domains to improve the recommendation quality," in *Proceedings of the KDD 2014*, pp. 801–810, New York, NY, USA, 2014.

[52] M. Tong, H. Shen, X. Jin, and X. Cheng, "Cross-domain recommendation: an embedding and mapping approach," in *Proceedings of the IJCAI 2017*, pp. 2464–2470, Melbourne, Australia, 2017.

[53] H. Lee, Im Jinbae, S. Jang, H. Cho, and S. Chung, "MeLU: meta-learned user preference estimator for cold-start recommendation," in *Proceedings of the KDD 2019*, pp. 1073–1082, Anchorage, AK, USA, 2019.

[54] M. Vartak, A. Thiagarajan, C. Miranda, J. Bratman, and H. Larochelle, "A meta-learning perspective on cold-start recommendations for items," 2017.

[55] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the ICML*, pp. 1126–1135, Sydney, Australia, August 2017.

Hindawi
Complexity
Volume 2020, Article ID 7975952, 8 pages
https://doi.org/10.1155/2020/7975952

Research Article

Maintenance and Operation Optimization Algorithm of PV Plants under Multiconstraint Conditions

Chi Hua,[1,2] Liang Kuang,[1] and Dechang Pi[2]

[1]*College of Internet of Things Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, Jiangsu, China*
[2]*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China*

Correspondence should be addressed to Dechang Pi; dc.pi@nuaa.edu.cn

Received 26 June 2020; Revised 9 August 2020; Accepted 5 October 2020; Published 28 October 2020

Academic Editor: Chuan Zhou

With the rapid increase in the photovoltaic (PV) plants, the real-time operation and maintenance of photovoltaic power generation equipment is very important. The maintenance and dispatching of decentralized power stations is still one of the key issues affecting the operation safety of photovoltaic power stations. However, most of the photovoltaic power stations in China fail to rationally optimize the utilization of resources and time. The current study puts forward effort implementation via genetic algorithm-based multiconstrained optimization methodology. The proposed study optimally overrides the traditional PV plant operation and maintenance dispatching operations with automation and reliability. The proposed study is also applicable to multiple PV plants, multiple maintainers, multipoint departure, different dispatching conditions, and cost considerations. We propose an MOOA algorithm to solve this issue, and we strongly believe that, by defining a suitable fitness function, the convergence speed and optimization ability can be greatly improved, and this study puts a forward step.

1. Introduction

Photovoltaic (PV) power plants are the core component of the smart grid [1, 2]. Because of its dispersed geographical location, complex structure, large number of equipment, and t complicated daily operation and maintenance procedures [3, 4], all these routine operations greatly affect the smooth operation of the smart grid. At present, the maintenance of PV plants [5–7] is mainly carried out in accordance with the regular maintenance stipulated by the state. There exist a number of problems which lead to excessive cost maintenance, such as insufficient maintenance, poor reliability, wasting of resources, and increasing maintenance costs. Therefore, operation and maintenance dispatch has become the core problem of an intelligent operation system of PV plants [8, 9]. Operation and maintenance assignment is essentially an extension of the assignment problem. The assignment of n tasks to n persons belongs to balanced assignment; otherwise, it belongs to an unbalanced assignment [10]. In the process of operation and maintenance of PV plants, the number of maintainers is less than that of power plants, which leads to the unbalanced assignment problem. Balanced assignment problem can be well solved by the Hungarian algorithm, eliminating heights and shrinking matrix analysis [11], while unbalanced assignment problem is an objective function optimization problem with constraints, and constraints and objective functions may not be linearized, so it is difficult to solve the problem. A couple of research studies have put forward efforts to solve an unbalanced assignment problem [12, 13], such as dual transportation method [14–17] and fuzzy Hungarian algorithm [18]. However, the solving process is complex and does not have robustness and adaptability. The machine learning especially deep learning also provides a solution to the unbalanced assignment problem [19]. In this paper, a genetic algorithm is presented to solve the unbalanced assignment problem of operation and maintenance dispatch. This method can optimize the operation and

maintenance dispatch of multiple PV plants, multiple maintainers, and multipoint departure. It can set different assignment constraints or cost considerations and accelerate the convergence speed and improve the optimization ability of the algorithm.

2. Establishment of Operational and Maintenance Model of PV Plants

The assignment problem takes the matching of tasks and maintainers as the research object to optimize the utilization of resources [20, 21]. The mathematical model of operation and maintenance dispatch for PV plants is described as follows.

Assuming that the number of maintainers is m and the number of tasks is n at a certain time, it obtains $m \le n$ by considering the maintenance cost, and one plant is only maintained by one maintainer; the maintainers will perform the next task according to the algorithm hints after the accomplishment of last PV plant's maintenance. In this paper, the optimal path of operation and maintenance dispatch is given to minimize the time spent by maintainers on the road. c_{ij} represents the time consuming on the road between two maintained PV plants, i.e.,

$$c_{ij} \ge 0, \quad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n. \tag{1}$$

Then, the mathematical model of multiobjective unbalanced assignment for m maintainers to n PV plants is as follows:

$$\min S_k = \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} x_{ij}, \quad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, p, \tag{2}$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} x_{ij} = 1, i = 1, 2, \ldots, m, \\ \sum_{j=1}^{n} \sum_{i=1}^{m} x_{ij} = n, \quad j = 1, 2, \ldots, n, \\ x_{ij} = 0 \text{ or } 1, \quad i = 1, 2, \ldots, m; j = 1, 2, \ldots, n, \end{cases} \tag{3}$$

where $x_{ij} = 1$ means that the i maintainer completes the j task and $x_{ij} = 0$ means that no task is assigned.

In order to solve the unbalanced assignment in (2), it makes an equivalent exchange for (2), that is, $n - m$ maintainers are added to make up m maintainers. In essence, the added maintainers are only considered in the calculation and will not be shown in dispatch route planning; moreover, the added maintainers will not affect the objective function and constraints, so the essence of the unbalanced assignment problem has not changed. The original unbalanced assignment problem is equivalent to the balanced assignment problem, that is, m maintainers are responsible for the operation and maintenance of n PV plants. The mathematical model is as follows:

$$\min S_k = \sum_{j=1}^{n} \sum_{i=1}^{n} c_{ij} x_{ij}, \quad i = 1, 2, \ldots, n; j = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, p, \tag{4}$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots, n, \\ \sum_{j=1}^{n} \sum_{i=1}^{n} x_{ij} = 1, \quad j = 1, 2, \ldots, n, \\ x_{ij} = 0 \text{ or } 1, \quad i = 1, 2, \cdots, n; j = 1, 2, \ldots, n, \end{cases} \tag{5}$$

where $x_{m+1,1}, x_{m+1,2}, \ldots x_{m+1,n}, \ldots, x_{n,n}$ attributing 0-1 variables correspond to the task assignment of adding maintainers. Obviously, the transformed model is equivalent to the original model. Equation (4) only corresponds to one assignment, which satisfies both constraints and objective optimization. Traditional methods are difficult to solve it. Essentially, the assignment problem of the operation and maintenance of PV plants is an NP-complete problem, which is solved by searching one by one, and time consumed will be $O(n!)$. In this paper, we have employed a genetic algorithm to solve the operation and maintenance dispatch of PV plants.

3. Optimization Algorithm of Maintainers Dispatching

The bases of the genetic algorithm for solving the optimal maintenance strategy of PV plants are as follows: First, the genetic algorithm solves various TSP problems with many results [22–24] and is mature and reliable. Second, the genetic algorithm satisfies the solution of different problems; for instances, the problem of multipoint departure and multiperson operation and maintenance can be assigned by adjusting the chromosome coding method, and the operation and maintenance strategy with uncertain time constraints and cost calculation methods can be made by adjusting the fitness function. Third, the genetic algorithm is easy to iterate, which meets the new demands emerging from the operation and maintenance environment of actual PV plants [25, 26].

As shown in Figure 1, the fault diagnosis process of PV array is described. This paper designs the optimal distribution of voltage sensors according to the distribution of PV plants by equivalent analysis of the weight matrix. Through regression analysis of state data [27, 28], the location of the fault point is obtained by analyzing the voltage signal detected via the voltage sensor. The fault diagnosis model of PV arrays is designed and carried out. This model classifies the types of fault points [29–31]. Based on the concept of equipment lifecycle management [32, 33], this paper establishes the equipment status data model, obtains the information of equipment management, establishes the health history of equipment status, refers to the provisions of equipment manufacturers on their design life, effective operation life, maximum allowable life, and the sample

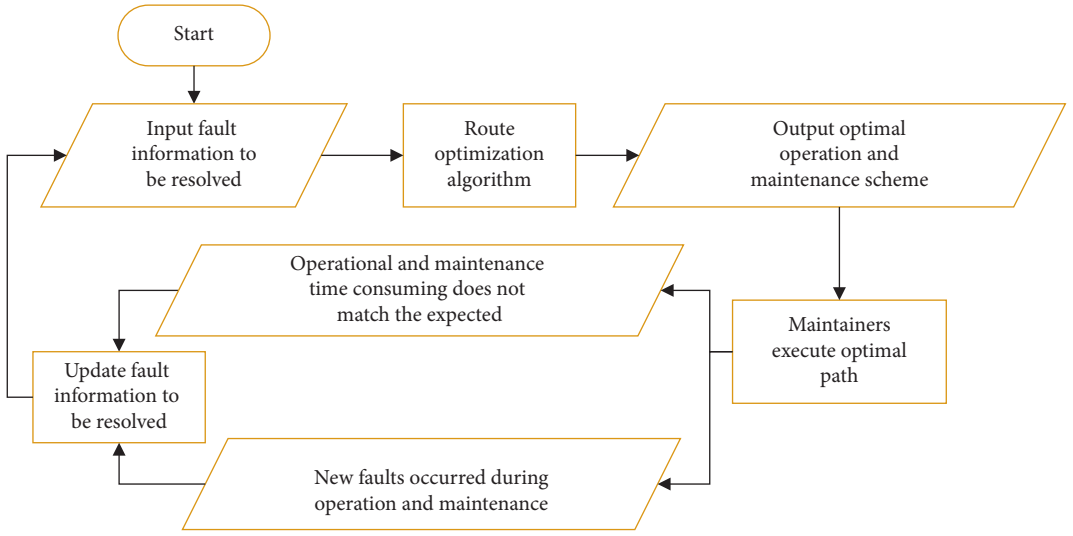


Figure 1: Operation process of fault diagnosis for PV arrays.

statistical data of their actual service life, and achieves the assistant decision making of equipment condition monitoring in photovoltaic power plants.

The following steps are used to optimize the operation and maintenance dispatch of PV plants:

Step 1. Input the fault information to be solved, including the number of faults l, the location of the faulty power station k, the installed capacity of the faulty power station c, the fault grade of the faulty power station g, the remaining maintenance waiting time of the faulty power station t, and the fault flow number s.

Step 2. The path optimization algorithm is obtained by the genetic algorithm.

① Input the traffic cost matrix, that is, the traffic cost between the fault location and the departure location.

② Random sequencing of fault pipeline numbers constitutes the initial gene as follows:

$$\text{gene} = [\, s_{a1},\ s_{a2},\ \cdots,\ s_{an} \,]. \tag{6}$$

③ Repeat ② for N times to form N initial genes: $\text{gene}_1, \text{gene}_2, \ldots, \text{gene}_n$; N initial genes constitute the initial population Q.

④ Construct fitness function $\Psi(\text{gene})$:

$$\Psi(\text{gene}) = \frac{a}{\sum_{i=1}^{n-1} C_{s_i, s_{i+1}}} + \frac{b}{\sum_{i=1}^{n} I(s_i) * c_{s_i}} + \frac{c}{\sum_{i=1}^{n} I(s_i) * g_{s_i}} + \frac{d}{\sum_{i=1}^{n} I(s_i) * t_{s_i}}, \tag{7}$$

where gene represents the gene of fitness function to evaluate fitness, s_i represents the fault flow number in the gene, $C_{s_i, s_{i+1}}$ represents the traffic cost between the two fault locations determined by the fault flow number and the traffic cost matrix C, $\sum_{i=1}^{n-1} C_{s_i, s_{i+1}}$ represents the total traffic cost of maintenance in the order of the intermediate flow number, $I(s_i)$ represents the order of s_i in gene; $I(s_i) = 1$, if it ranks first in gene. c_{s_i} denotes the installed capacity of the faulty power station corresponding to the fault pipeline number s_i, g_{s_i} denotes the fault grade of the faulty plant corresponding to the fault pipeline number s_i, and t_{s_i} denotes the remaining maintenance waiting time of the faulty power station corresponding to the fault pipeline number s_i. a represents the penalty coefficient of the traffic cost; the default is 1000; the larger m is, the lower the route of the traffic cost will be preferred, and b represents the penalty coefficient of the installed capacity; the default is 100; the larger b is, the preferred choice will be made. For a power station with large installed capacity, c means the penalty coefficient of the fault grade. The bigger c is, the higher the penalty coefficient of maintenance time will be. d means the penalty coefficient of maintenance time. The bigger d is, the lower the waiting time of maintenance will be.

⑤ n_1 individuals were randomly selected from the population Q, and the probability of individual gene_i being selected was as follows:

$$P_i = \frac{\Psi(\text{gene}_i)}{\text{sum}(\Psi(\text{gene}))}, \tag{8}$$

that is, the greater the fitness, the higher the probability of being selected, and the new individuals selected are inserted into the new population Q'.

⑥ The probability P_c is used to randomly select n_2 individuals from the population Q, which is even and convenient for crossover operation. The crossover operation is carried out by randomly extracting a number of individual gene fragments and pairing individuals to form a new individual to join the population Q'. In the optimization process

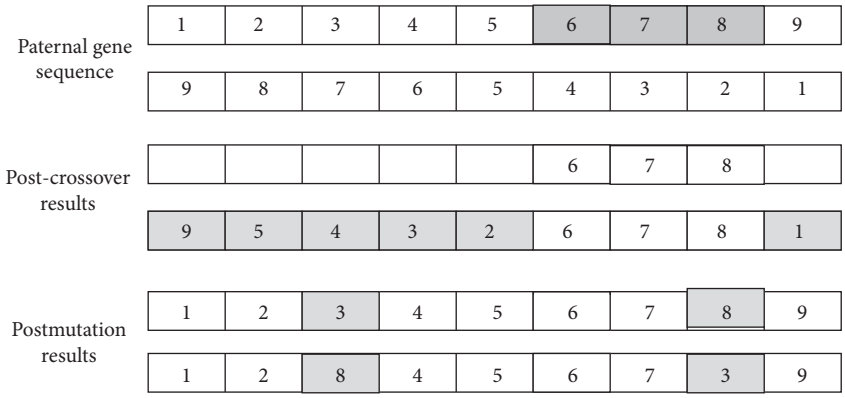


Figure 2: Gene crossover and mutation.

of operation and maintenance of PV plants, P_c is set to 0.75.

In the new population Q', the probability P_m is used to randomly select n_3 individuals, which is for mutation. That is to say, two gene fragments are randomly selected, and their positions are exchanged to form n_3 new individuals to join the population Q'. In the optimization process of operation and maintenance of PV plants, P_m is set to 0.01. The crossover method and the mutation method are shown in Figure 2.

Two individuals were selected randomly with probability P_c to perform crossover operation. The offspring inherited part of the genes from their parents and kept the same sequence as their parents are as shown in Figure 2. Some genes of Parent 1, such as 678, crossed with Parent 2 were selected randomly, and the crossover results are shown in Figure 2.

The probability P_m of mutation occurs. Because each city only passes through once in TSP problem, the mutation should not change the value of one position in the gene sequence (which will cause a city to pass through twice) but should exchange the values of two positions randomly, as shown in Figure 2; the positions of 3 and 8 should be exchanged.

⑦ Repeat ⑤ and ⑥ until the maximum number of iterations is reached, or the fitness function of an individual reaches a given threshold.

⑧ For the individual with the largest output fitness, the sequence of flow numbers is the maintenance sequence given by the algorithm.

Step 3. The optimal operation and maintenance scheme is output by the path optimization algorithm.

Step 4. If when the maintainers implement the optimal path, the operation and maintenance time consuming does not match the expected time consuming or new faults occur during the operation and maintenance period; we need to update the information to solve the fault and repeat Steps 2–4 again, until the maintainers successfully implement the optimal path.

4. Experiments and Implementation Verification

The experimental environment uses MATLAB 6.0 to construct 36 random points as the distribution of PV plants in the region and triangulate the points, and the time spent on the road is the sum of the edge lengths of the triangle experienced between the two points; construct the weight matrix of the operation and maintenance process, randomly select a number of points from the points as maintenance objects, and set them up again every 1 minute from the points. Randomly select 0 to 1 point as a new maintenance object to join the maintenance queue. Suppose a maintainer is at a fixed speed between the set of points in the operation and maintenance, and the speed per minute does not exceed the length of the maximum edge of the triangle. As shown in Figure 3, randomly generate points, triangulate points, and construct maintenance site.

In the experiment, the maintenance of PV plants can be well carried out by implementing one maintainer and three maintainers, respectively. By implementing a genetic algorithm, the optimal dispatching path of maintainers can be saved, especially when facing 38 photovoltaic power stations, no more than three maintainers can be maintained.

In order to better prove the rationality of the optimization method of operation and maintenance dispatch of PV plants based on genetic algorithm, random method, greedy algorithm, and genetic algorithm were used to carry out experiments. The time consuming on the road between two places of a company's PV plants in Zhejiang province is shown in Table 1, including Hangzhou, Deqing, Haining, Huzhou, Jiande, Kaihua, Kecheng, Linan, Linhai, Longquan, Pan'an, Quzhou, Shangyu, Shaoxing, Shengzhou, Taizhou, Tiantai, Tonglu, Tongxiang, Yuyao, and Zhuji. The PV plants in all areas are repaired and maintained once to find an appropriate operation and maintenance path. The average travel time for each region is 159 minutes, of which

(1) The average total time of 21 horizons patrol by the random route method is 3340 minutes.

(2) Using a greedy algorithm (the next destination is the nearest location from the current location), the route is selected as follows:

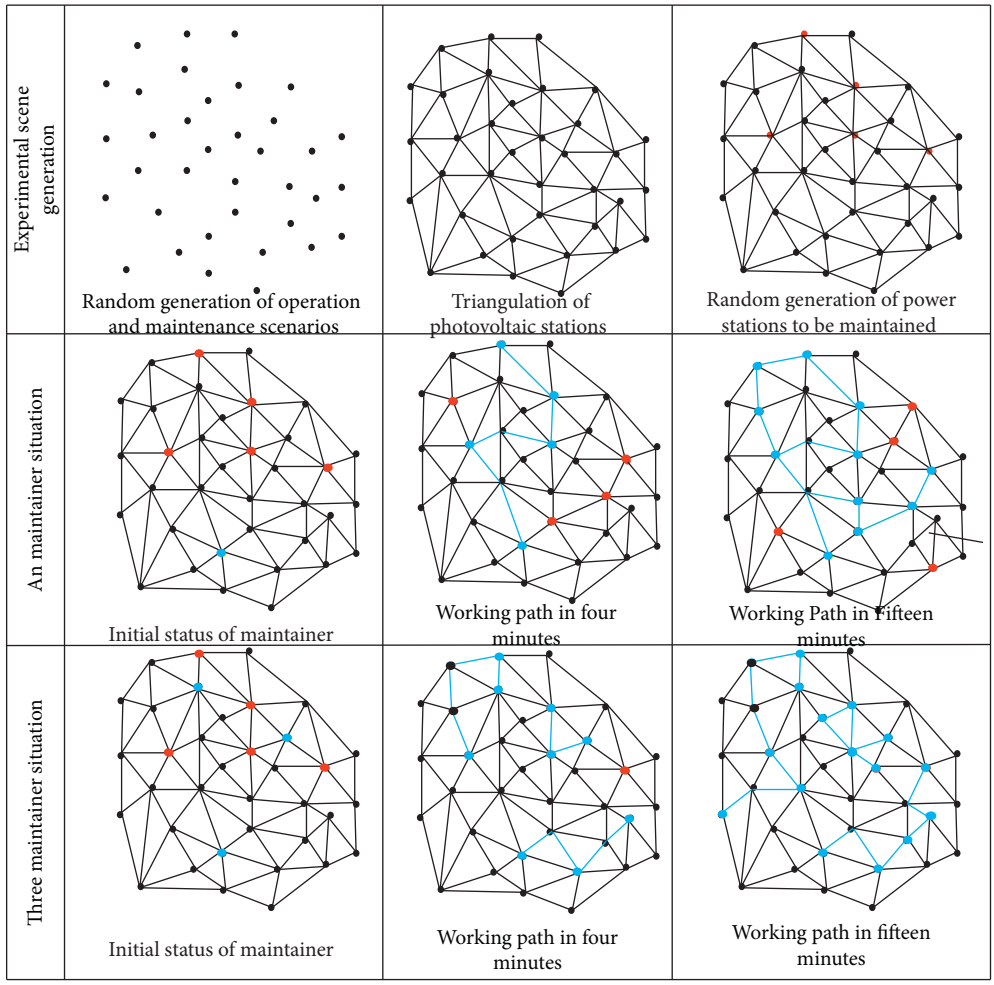


Figure 3: Scenario construction and operation and maintenance path optimization.

Hangzhou > Tongxiang > Haining > Shangyu > Shaoxing > Yuyao > Shengzhou > Tiantai > Linhai > Taizhou > Pan'an > Zhuji > Linan > Deqing > Huzhou > Tonglu > Jiande > Kecheng > Quzhou > Kaihua > Longquan > Hangzhou.

The time series is {61, 35, 65, 45, 40, 86, 69, 54, 68, 144, 99, 98, 73, 58, 140, 66, 71, 13, 61, 217, 313}, totaling 1876 minutes.

(3) The genetic algorithm is used to optimize the maintenance path. The time cost matrix of operation and maintenance location is shown in Table 1. The selected route is as follows:

Hangzhou > Linan > Deqing > Huzhou > Tongxiang > Haining > Zhuji > Shaoxing > Yuyao > Shangyu > Shengzhou > Tiantai > Taizhou > Linhai > Pan'an > Longquan > Kaihua > Kecheng > Quzhou > Jiande > Tonglu > Hangzhou.

The time series is {62, 73, 58, 87, 35, 103, 77, 40, 55, 67, 69, 90, 68, 97, 187, 217, 56, 13, 76, 66, 78}, totaling 1278 minutes.

Compared with the optimal path obtained by the exhaustive wet method, it is confirmed that the improved genetic algorithm can obtain the optimal path for 21 operation and maintenance routes.

The coding method of the genetic algorithm and the definition of fitness function in [34, 35] were used to solve the operation and maintenance route planning of PV plants. Comparing with this algorithm, as shown in Figure 4, [35] needs 81 iterations, [34] needs 44 iterations, while this algorithm only needs 34 iterations, and the final operation and maintenance path takes 1278 minutes. By defining an appropriate fitness function, the strategy of a genetic algorithm to optimize the operation and maintenance path presented in this paper can accelerate the convergence speed and searchability of the genetic algorithm.

5. Summary

This paper presents an optimization method of operation and maintenance dispatch of PV plants based on the genetic algorithm. This method aims at the operation and maintenance tasks of large-scale distributed PV plants. GA

Table 1: Time cost matrix of operation and maintenance sites (unit: minutes).

		0 Hangzhou	1 Deqing	2 Haining	3 Huzhou	4 Jiande	5 Kaihua	6 Kecheng	7 Linan	8 Linhai	9 Longquan	10 Pan'an	11 Quzhou	12 Shangyu	13 Shaoxing	14 Shengzhou	15 Taizhou	16 Tiantai	17 Tonglu	18 Tongxiang	19 Yuyao	20 Zhuji
0	Hangzhou	0	68	67	96	114	198	166	62	189	313	154	170	71	70	104	221	144	78	61	88	82
1	Deqing	68	0	83	58	143	226	194	73	216	339	176	209	101	100	132	248	170	110	92	115	110
2	Haining	67	83	0	98	154	238	204	114	176	313	162	209	65	83	94	210	132	119	35	79	103
3	Huzhou	96	58	98	0	175	265	230	119	245	370	206	237	131	132	160	283	200	140	87	146	139
4	Jiande	114	143	154	175	0	102	71	125	219	210	170	76	149	139	177	258	213	66	169	165	136
5	Kaihua	198	226	238	265	102	0	56	197	241	217	191	61	227	216	199	281	235	135	250	254	166
6	Kecheng	166	194	204	230	71	56	0	182	197	189	150	13	188	177	156	236	194	100	214	207	127
7	Linan	62	73	114	119	125	197	182	0	244	315	167	176	129	105	164	281	200	89	104	144	98
8	Linhai	189	216	176	245	219	241	197	244	0	228	97	208	148	167	112	68	54	261	205	151	172
9	Longquan	313	339	313	370	210	217	189	315	228	0	187	192	280	281	241	276	267	247	331	313	255
10	Pan'an	154	176	162	206	170	191	150	167	97	187	0	164	132	127	88	144	88	171	175	156	99
11	Quzhou	170	209	209	237	76	61	13	176	208	192	164	0	195	187	173	248	210	108	226	213	136
12	Shangyu	71	101	65	131	149	227	188	129	148	280	132	195	0	45	67	186	117	120	99	55	87
13	Shaoxing	70	100	83	132	139	216	177	105	167	281	127	187	45	0	68	187	109	120	99	40	77
14	Shengzhou	104	132	94	160	177	199	156	164	112	241	88	173	67	68	0	153	69	154	117	86	107
15	Taizhou	221	248	210	283	258	281	236	281	68	276	144	248	186	187	153	0	90	273	234	179	212
16	Tiantai	144	170	132	200	213	235	194	200	54	267	88	210	117	109	69	90	0	191	161	127	144
17	Tonglu	78	110	119	140	66	135	100	89	261	247	171	108	120	120	154	273	191	0	121	136	107
18	Tongxiang	61	92	35	87	169	250	214	104	205	331	175	226	99	99	117	234	161	121	0	101	105
19	Yuyao	88	115	79	146	165	254	207	144	151	313	156	213	55	40	86	179	127	136	101	0	97
20	Zhuji	82	110	103	139	136	166	127	98	172	255	99	136	87	77	107	212	144	107	105	97	0

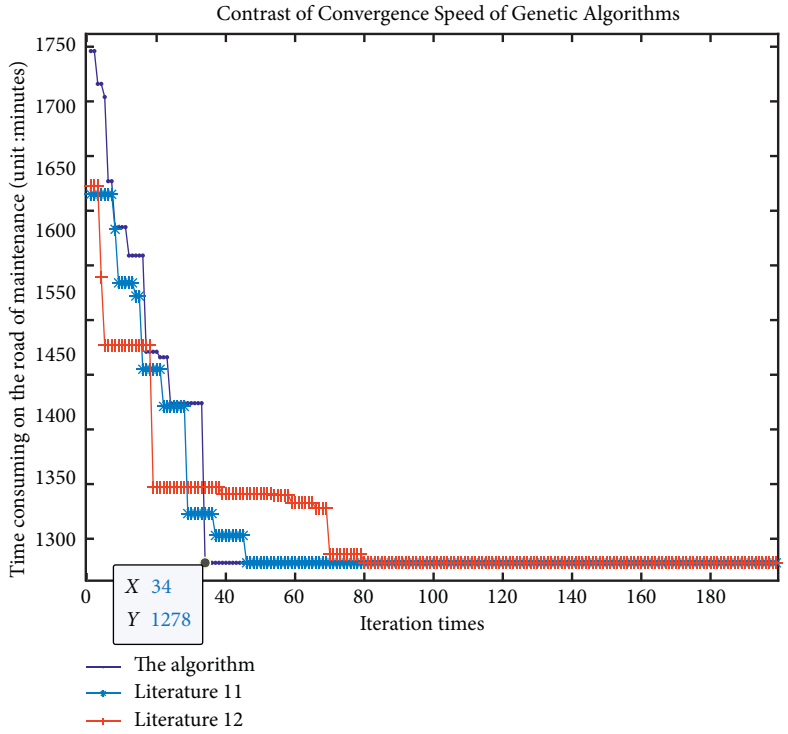


Figure 4: Comparison of convergence speed of genetic algorithms.

for M-TSP is applied to provide an optimal path for maintainers dispatched for PV plants. It supports the optimization of operation and maintenance dispatch of multiple PV plants, multimaintainer, multiobjective, and multipoint departure. By multitype cost function, different assignment constraints or cost considerations can be set, such as arrival within a specified time or minimum time cost of transportation costs. We ascertain that the proposed model will lead the way.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

Some of the authors of this publication are also working on these related projects: (1) Higher Vocational Education Teaching Fusion Production Integration Platform Construction Projects of Jiangsu Province under grant no. 2019(26), (2) Natural Science Fund of Jiangsu Province under grant no. BK20131097, (3) "Qing Lan Project" Teaching Team in Colleges and Universities of Jiangsu Province under grant no. 2017(15), (4) High Level of Jiangsu Province Key Construction Project funding under grant no. 2017(17), and (5) Natural Science Research in Jiangsu Universities in 2020.

References

[1] S. Prakesh, S. Sherine, and B. Bist, "Forecasting methodologies of solar resource and PV power for smart grid energy management," *International Journal of Pure and Applied Mathematics*, vol. 116, no. 18, pp. 313–318, 2017.

[2] M. S. Hossain, N. A. Madlool, N. A. Rahim, J. Selvaraj, A. K. Pandey, and A. F. Khan, "Role of smart grid in renewable energy: an overview," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 1168–1184, 2016.

[3] S. S. S. R. Selvaraj, L. Wang, V. Devabhaktuni, and N. Gudi, "Smart meters for power grid—challenges, issues, advantages and status," in *Proceedings of the IEEE/PES Power Systems Conference and Exposition*, pp. 1–7, Phoenix, AZ, USA, March 2011.

[4] Y. Kabalci, "A survey on smart metering and smart grid communication," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 302–318, 2016.

[5] M. De Benedetti, F. Leonardi, F. Messina, C. Santoro, and A. Vasilakos, "Anomaly detection and predictive maintenance for photovoltaic systems," *Neurocomputing*, vol. 310, no. 1, pp. 59–68, 2018.

[6] V. Mauro, C. Vittorio, L. Alfonsi, and V. Introna, "Optimization of photovoltaic maintenance plan by means of a FMEA approach based on real data," *Energy Conversion and Management*, vol. 152, no. 1, pp. 1–12, 2017.

[7] M. Di Somma, B. Yan, N. Bianco et al., "Multi-objective design optimization of distributed-energy systems through

cost and exergy assessments," *Applied Energy*, vol. 204, no. 1, pp. 1299–1316, 2017.

[8] S. Affijulla and S. Chauhan, "A new intelligence solution for power system economic load dispatch," in *Proceedings of the 10th International Conference on Environment and Electrical Engineering*, pp. 1–5, Rome, Italy, May 2011.

[9] J. Gao, Y. Xiao, J. Liu, W. Liang, and C. L. P. Chen, "A survey of communication/networking in Smart Grids," *Future Generation Computer Systems*, vol. 28, no. 2, pp. 391–404, 2012.

[10] W. Liang, D. Pi, and Z. Shao, "A pareto-based estimation of distribution algorithm for solving multiobjective distributed No-wait flow-shop scheduling problem with sequence-dependent setup time," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1344–1360, 2019.

[11] S. H. I. Zhong-min, "General assignment problem," *Operations Research and Management Science*, vol. 8, no. 1, pp. 21–26, 1999.

[12] Y. Song, M.-Y. Chen, and Z. Shu-hong, "Multi-objective decision-making method for site selecting of projects based on fuzzy assignment," *System Engineering—Theory and Practice*, vol. 22, no. 1, pp. 108–111, 2002.

[13] H. De-cai, "An efficiency algorithm for solving the optimal solution of a generalized assignment problem," *Control and Decision*, vol. 14, no. 3, pp. 272–275, 1999.

[14] F. Glover, D. Karney, and D. Klingman, "A computation study on start procedures, basis change criteria, and solution algorithms for transportation problems," *Management Science*, vol. 20, no. 5, pp. 793–813, 1974.

[15] S. Napier, "General assignment problem," *Operations Research and Management Science*, vol. 1, 1999.

[16] A. Kumar, S. S. Appadoo, and P. Kaur, "Mehar approach for solving dual-hesitant fuzzy transportation problem with restrictions," *Sādhanā*, vol. 45, no. 1, pp. 1–9, 2020.

[17] L. Cui, J. Wu, D. Pi, P. Zhang, and P. Kennedy, "Dual implicit mining-based latent friend recommendation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 5, pp. 1663–1678, 2018.

[18] X.-Z. Qin, W. Xue-hua, "Dynamic programming model of a sort of optimal assignment problem," *Mathematics in Practice and Theory*, vol. 26, no. 3, pp. 212–216, 1996.

[19] F. Liu, S. Xue, J. Wu et al., "Deep Learning for Community Detection: Progress, Challenges and Opportunities," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, Yokoham, Japan, July 2020.

[20] B. Ming, P. Liu, S. Guo, X. Zhang, M. Feng, and X. Wang, "Optimizing utility-scale photovoltaic power generation for integration into a hydropower reservoir by incorporating long- and short-term operational decisions," *Applied Energy*, vol. 204, pp. 432–445, 2017.

[21] F.-F. Zhang and J. Qiu, "Multi-objective optimization for integrated hydro-photovoltaic power system," *Applied Energy*, vol. 167, pp. 377–384, 2016.

[22] J. Yan-Qun, Z. Shu-Guang, and C. Bokui, "Fixed-point fast sweeping weighted essentially non-oscillatory method for multi-commodity continuum traffic equilibrium assignment problem," *Applied Mathematical Modelling*, vol. 62, pp. 404–414, 2018.

[23] Z. Wang, P. Jun, C. Liling et al., "A parallel biological optimization algorithm to solve the unbalanced assignment problem based on DNA molecular computing," *International Journal of Molecular Sciences*, vol. 16, no. 10, pp. 25338–25352, 2015.

[24] S.-U. Lee, "Assignment problem algorithm using the dual-path competition method," *Journal of Korean Institute of Information Technology*, vol. 13, no. 12, pp. 75–83, 2015.

[25] P. A. Kumari and P. Geethanjali, "Parameter estimation for photovoltaic system under normal and partial shading conditions: a survey," *Renewable and Sustainable Energy Reviews*, vol. 84, pp. 1–11, 2018.

[26] T. S. Babu, J. P. Ram, T. Dragičević et al., "Particle swarm optimization based solar PV array reconfiguration of the maximum power extraction under partial shading conditions," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, pp. 74–85, 2018.

[27] C. A. MacKenzie and K. Barker, "Empirical data and regression analysis for estimation of infrastructure resilience with application to electric power outages," *Journal of Infrastructure Systems*, vol. 19, no. 1, pp. 25–35, 2013.

[28] M. Hassanzadeh and C. Y. Evrenosoglu, "A regression analysis based state transition model for power system dynamic state estimation," in *Proceedings of the 2011 North American Power Symposium*, Boston, MA, USA, August 2011.

[29] J. Wu, S. Pan, X. Zhu et al., "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2014.

[30] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[31] C. Yu, D. Pi, S. I. Becker, L. Cui, and B. Johnson, "CenEEGs," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 2, pp. 1–25, 2020.

[32] D. Wu, A. Bufardi, and P. Xirouchakis, "Research issues on product lifecycle management and information tracking using smart embedded systems," *Advanced Engineering Informatics*, vol. 17, no. 3-4, pp. 189–202, 2003.

[33] L. A. I. Yun and L. U. Chen, "Construction of entire lifecycle management model of laboratory equipment in universities," *Research and Exploration in Laboratory*, vol. 2, 2012.

[34] I. Younas, F. Kamrani, M. Bashir et al., "Efficient genetic algorithms for optimal assignment of tasks to teams of agents," *Neurocomputing*, vol. 314, pp. 409–428, 2018.

[35] B. Marin, D. Guillerme, and G. Rodolphe, "Approximability and exact resolution of the multidimensional binary vector assignment problem," *Journal of Combinatorial Optimization*, vol. 36, no. 3, pp. 1059–1073, 2018.

Hindawi
Complexity
Volume 2020, Article ID 5346085, 8 pages
https://doi.org/10.1155/2020/5346085

WILEY | Hindawi

Research Article

A Novel Expert Finding System for Community Question Answering

Nan Zhao,[1] **Jia Cheng**,[1] **Nan Chen**,[1] **Fei Xiong**,[2] **and Peng Cheng**[3]

[1]*State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China*
[2]*Key Laboratory of Communication & Information Systems, Beijing Jiaotong University, Beijing 100044, China*
[3]*Space Star Aerospace Technology Applications Limited Corporation, Xi'an 710077, China*

Correspondence should be addressed to Nan Zhao; zhaonan@xidian.edu.cn

Received 23 June 2020; Revised 10 August 2020; Accepted 6 October 2020; Published 23 October 2020

Academic Editor: Jia Wu

With the popularity of community question answering (CQA) sites, the research on identifying the expert users in online communities attracted increasing attention. We present a novel expert ranking algorithm based on the quality of user posts and the authority of user in community, and the similarity between the knowledge tags of users and questions in CQA sites is adopted in our scheme. Experimental results show that our scheme has better performance and accuracy under the same background with an amount of data samples.

1. Introduction

The number of Internet users is growing rapidly, along with the fast development of the network applications and infrastructure. In the enjoyment of convenience the network brings, it becomes difficult for the users to obtain the effective acquisition and screening of information [1, 2]. It follows then that community question answering (CQA) sites spring up [3, 4]. CQA sites are online knowledge communities, specializing in knowledge sharing and seeking, such as Stack Overflow [5, 6] and Yahoo Answers [7]. CQA sites provide a network platform for users to ask and answer questions and achieve information transfer and knowledge sharing among Internet users. Due to various topics and abundance content in CQA sites, network users prefer CQA sites to conventional web pages when seeking topic-specific information or solving problems [8]. The quality of information provided by CQA sites has been greatly improved in recent years. However, with the increasing number of community users, online communities amass an enormous amount of knowledge, which contains many useless answers inevitably in the community. Therefore, it is crucial to identify and recommend the experts in different fields of the CQA sites for the community operation and extension services [9–11]. Meanwhile, network users can gain the accurate and high quality experience. Therefore, expert finding technique is of significance to improve the accuracy and efficiency of information acquisition in the CQA sites [12–14].

The existing expert finding techniques [15, 16] are divided into three major categories in generally. A directed graph is built based on the interaction between network users in the community, and the users are ranked by adopting the link analysis algorithm [17, 18] in the first category. In the second category, the text data in the community is analyzed based on the topic models [19, 20], and the results are applied to expert recommendation [21, 22]. In the last category, the hybrid models are built for expert finding with the methods mentioned above. A number of strategies are proposed unceasingly, but there are still some imperfections. Most of the traditional expert finding techniques ignore the user's activeness in the community. It may lead to the expert users not providing timely response. In addition, the comprehensive factors are not considered completely in some methods when users' expertise is evaluated. Finally, it may lead to the limited authority of recommended experts.

In our study, a more complete expert finding system that includes the expert ranking and the expert recommendation

is proposed. A new expert ranking algorithm is presented in this paper, named as Exp-rank. Exp-rank considers not only the authority of users in the community but also the quality of content published by users. On the basis of expert ranking, we calculate the similarity between the new questions and the knowledge tags of the users. According to the calculated results, we recommend experts more accurately to the new question. The rest of the paper is organized as follows. Section 2 briefly introduces the related work. Section 3 presents the proposed Exp-rank algorithm and the expert recommendation. Section 4 provides the dataset, performance evaluation, results, and discussion. Finally, section 5 summarizes the full task.

2. Related Work

The expert finding in online community is a widely investigated problem [23, 24]. In the development of CQA sites, a large number of users register for the community and participate in topic discussions. Meanwhile, the community has accumulated a lot of content, including a lot of useless information. Expert finding techniques can help identify and recommend the expert users in the CQA sites and avoid the adverse effects caused by spam and useless information [25–27]. The results of expert finding can be applied to the information management of the CQA sites and are helpful to provide users with more efficient and accurate question and answer service.

Link analysis algorithm is significantly adopted in the research of the expert ranking [28]. Zhang et al. [29] propose and evaluate several link-based expert ranking algorithms. They reveal that PageRank-based expert ranking algorithms outperform other algorithms in the online community. Yang and Wu [30] adopt weighted HITS algorithm to find experts in CQA sites. Link analysis algorithm can reflect the authority of users in the community. However, the link-based expert finding techniques focus merely on the link structure among individuals. They ignore the impact of useless replies and advertising accounts.

Graph-based algorithms are also applied in the research of expert finding. Zhao et al. [31] consider the problem of expert finding from the viewpoint of missing value estimation. The performance of the expert finding in CQA systems is improved because they employ users' social networks for inferring user model. Aslay et al. [32] propose Competition-Based Expertise Networks (CBEN), a novel community expertise network structure based on the principle of competition among the answerers of a question.

On the other hand, some researches reveal experts by analyzing online community content and user profile. Because of the complexity and diversity of information, the related strategies are varied. Shao and Yan [33] propose a model with two prediction methods that include the traditional feature-based method and LDA method. Specifically, when a new question arises, the model adopts LDA method to label and classify the question according to the latent semantic and content features. Then, with the traditional features of the question and the asker information, the model can recommend the appropriate expert users to answer this new question. Lu et al. [34] use semantic information extracted from user interaction to identify expert users. They construct the user question-answer interaction graph through direct semantic links and potential links extracted from the records of question session records and user profiles. After that, they employ the semantic information in the propagation link analysis method and in the language model. Faisal et al. mainly adopt the reputation of users in the community and the quality of users' answers as the main experts' evaluation indexes. On this basis, they combine voter reputation, voting rate, and other characteristics to measure the user expertise [35–37]. However, not every CQA site provides users with services like reputation system. Therefore, it is not conducive to extend these researches to other network communities.

Expert finding techniques based on social network features are rare. Most studies take the features of social network as one of the indicators to evaluate experts and propose a hybrid model for expert finding [38]. Wang et al. [39] consider both the relevance of documents and the authority of users in the community to assess the level of experts. Rafiei and Kardan [40] propose a hybrid method for expert finding in online communities, which is about the content analysis and the social network analysis. The content analysis is based on the concept map and the social network analysis is based on PageRank algorithm. Zhou et al. [41] present a topic-sensitive probabilistic model, which is an extension of PageRank algorithm to find experts in CQA. Compared with the conventional link analysis technology, their method considers not only the link structure, but also the topic similarity between different users. In fact, most of the previous works focus only on the static ranking or matching of domain experts without considering comprehensive factors that influence the user's expertise. In particular, our work serves as a method of dynamic expert finding system that combines expert ranking and expert recommendation.

3. Materials and Methods

We propose a new expert finding system containing expert ranking and expert recommendation for CQA sites. We adopt the cumulative the quality factor and the authority of users as the expert evaluation indicators and recommend experts to the new questions. A generic overview of the proposed scheme is given in Figure 1. In CQA site, a group of questions are $Q = \{q_1, q_2, \ldots, q_n\}$, where the question q_i owns answers $A = \{a_1, a_2, \ldots, a_m\}$ from users $U = \{u_1, u_2, \ldots, u_i\}$. In particular, we identify the expert users based on the expertise and the authority from U. In more detail, we evaluate the expertise of the users by analyzing their past performance. Meanwhile, the link-based ranking algorithm is employed to calculate the authority of the users. Then, we combine the expertise and the authority of the users to identify expert users. For recommending experts more accurately for the new questions, we extract the knowledge tags from top-ranking experts, and obtain appropriate recommended expert users by calculating the

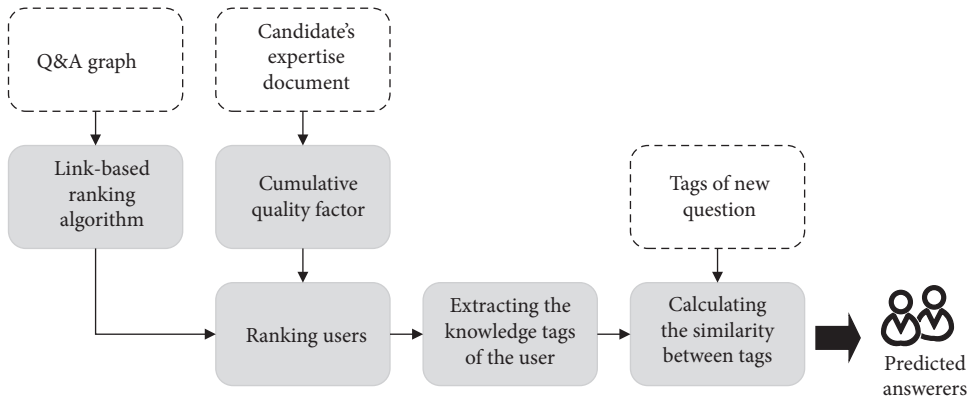


Figure 1: Architecture of proposed expert finding system.

similarity between the user knowledge tags and the new questions in the community.

3.1. Expert Authority Ranking Algorithm. We build a particular network that indicates the interaction of community members to determine the social influence of the users in the online community. We adopt Q&A graph to represent a social network based on the interaction of the users. In Q&A graph, the nodes represent different users in the community, and a directed edge is built between two users when they are the inquirer and the responder about the same question, respectively, as shown in Figure 2. Q&A graph describes the interactions of user in the online community. Link analysis algorithms include PageRank and HITS and the authority of nodes can be measured by link analysis algorithms based on Q&A graph. Lü et al. [42] proposed LeaderRank algorithm based on PageRank. As shown in Figure 3, they consider a network of N nodes and M directed links. Nodes correspond to users and links are established according to the relations among leaders and fans. The idea of LeaderRank algorithm is to add a ground node which connects to every user through bidirectional links (see Figure 3 for an illustration). The network thus becomes strongly connected and consists of $N+1$ nodes and $M+2N$ links. The out-degree or in-degree of all nodes is greater than zero, which avoids isolated nodes in complex networks and ensures the convergence of the algorithm. Moreover, LeaderRank algorithm is an adaptive parameter free algorithm. Comparing with PageRank, LeaderRank has higher accuracy and robustness in mining important network nodes.

Figure 2 illustrates the user relationship in the CQA sites. Nodes represent users and links are established according to the relations among inquirer and the responder. In the discussions of the online community, the question from u_1 is answered by u_2, can gain, u_2, a vote of support from u_1. If u_3 and u_4 answer the question from u_2, the vote of support from u_2 is evenly distributed to them. The expert authority ranking algorithm is based on the fact that a user owns more authority than the user whose question is answered by him.

In particular, the community has some user groups whose members rarely communicate with users outside the group, and these user groups usually contribute less to the mainstream topics in the community. However, more internal links may exist in these user groups. These internal links are worthful to improve the quality of ranking, but they are usually ignored in link analysis algorithms. Thus, we propose an expert authority ranking algorithm to measure the authority of the users based on the above. E consists of directed links formed by the relationship between question and answer from users in the community. We present u_{ji} to indicate the contribution of user i to user j:

$$u_{ji} = \begin{cases} 0, & \text{if } (j,i) \notin E, \\ n(j,i), & \text{if } (j,i) \in E, \end{cases} \tag{1}$$

$$u'_{ji} = u_{ji} - \beta u_{ij}, \tag{2}$$

$$a_{ji} = \frac{u'_{ji}}{\sum_{k=1}^{N} u'_{jk}}, \tag{3}$$

where $n(j,i)$ is the number of times that user i has answered j. β denotes a damping factor, and the range of β is from 0 to 1. The value of u'_{ji} equals the value that the backlinks between users are subtracted from u_{ji}. This helps to eliminate the effect of internal links in user groups on ranking. When $u'_{ji} < 0$, we set u'_{ji} equal to zero. N is the total number of users that have answered user j, and a_{ji} represents the vote of support from user j to user i. The authority value of user i at time t is $AU_i(t)$, and we have

$$AU_i(t+1) = \sum_{j=1}^{N+1} a_{ji} AU_j(t). \tag{4}$$

$AU_i(0) = 1$ represents the initial score of all user nodes except the ground node, and $AU_g(0) = 0$ represents the initial score of the ground node. AU_i tends to be stable at t_c, and the final authority score of user i is

$$AU_i = AU_i(t_c) + \frac{\theta_i AU_g(t_c)}{N}, \tag{5}$$

$$\theta_i = \theta * e^{-(d-d_0)}. \tag{6}$$

$AU_g(t_c)$ is the score of the ground node when it reaches the steady state. θ_i is the time attenuation of user i, θ is the attenuation coefficient, and d_0 represents the user's last post

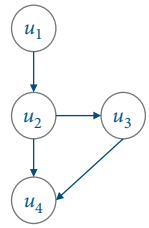


Figure 2: User-user relationship graph.

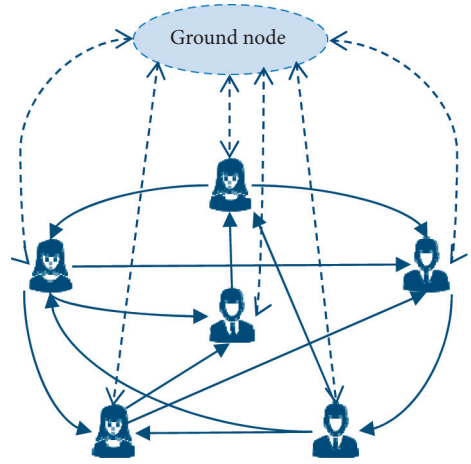


Figure 3: An illustration of the LeaderRank algorithm.

before the deadline d. If the last post from the user emerged a year ago, the attenuation coefficient of the user is 0. We measure the activity of the user according to the time attenuation and assign the score of the ground node through the activity of the users. Through the formula above, we can get the authority score of the users.

3.2. Cumulative Quality Factor. Most of the online communities possess malicious registered accounts, which are generally active in the communities, disseminating advertising or spam information. These accounts cannot be screened out when the authorities of users are calculated. Although some users actively participate in discussions to improve their authority value, the posts of these users are not professional and have little reference value for other users.

To solve these problems, we propose the concept of cumulative quality factor. In the process of data acquisition, we remove the users that rarely speak. Then, we summarize the scores or likes (the positive comments) of all the posts produced by users in the community and calculate the cumulative quality factor AS of the users:

$$AS_i = \frac{1}{N} \sum_{j=1}^{N} \ln\left(\delta_j + 1\right). \tag{7}$$

The total number of answers posted by user i is N. δ_j represents the score for the answer from user i, and $\delta_j + 1$ is to avoid the situation that the score is zero.

We can estimate the past performance of the users by calculating the cumulative quality factor. Moreover, it helps us to remove useless accounts and identify expert users with the expertise and the continuous excellent performance.

3.3. Exp-Rank. In order to evaluate expert users comprehensively, we combine the cumulative quality factor and the authorities of the users into an expert ranking standard. It is expressed as follows:

$$\mathrm{Exp}_i = AU_i \cdot AS_i^{\lambda}. \tag{8}$$

The expert score of user i is calculated with the cumulative quality factor AS_i and the authority value AU_i of the user. λ denotes a weighting factor, and the range of λ is from 0 to 1. In the expert finding of knowledge community, we think that the weight of AS_i should be slightly less than AU_i, and therefore the value of λ is set to 0.9. At last, we rank the candidates according to their expert scores Exp and obtain the results of expert ranking.

3.4. Expert Recommendation. We establish user files for the top ranking candidates. The user files are composed of questions and answers posted by users. The low-score answers posted by users are not adopted in the user files. Moreover, we extract keywords from user files by applying RAKE [43] algorithm. RAKE algorithm adopts punctuations to divide a file into several clauses, and the stop words are as delimiters to divide the clauses into several phrases, which are the candidates for the final extracted keywords. Each phrase can be split into several words by spaces, and every word can be given a score expressed as follows:

$$\text{word Score}(w) = \frac{\text{word Degree}(w)}{\text{word Frequency}(w)}. \quad (9)$$

In the formula, the value of word Degree (w) consists of two portions, which are the number of times the word w forms a phrase with other words and the total number of times the word w occurs in the file, and word Frequency (w) represents the total number of times the word w occurs in the file. The word Score (w) of each word is calculated by formula. The score of every phrase can be obtained by accumulating the scores of words. RAKE algorithm extracts phrases that score in the top third as keywords.

As shown in Table 1, we take the keywords of user files as the knowledge tags of the user. We employ cosine similarity to calculate the similarity between the knowledge tags of user and the tags of the new question because all the questions in Stack Exchange website possess their own tags [44]. The expert users are recommended to the new question according to the similarity scores.

4. Experiments

We adopt the dataset of Stack Exchange website to simulate and compare the results with other algorithms for verifying the effectiveness of our proposed method.

4.1. Dataset. Stack Overflow is an online knowledge community, originally designed for programmers and computer engineers. It was founded in 2008 by two programmers, Joel Spolsky and Jeff Atwood. Users can post and answer questions, discuss with each other, and retrieve information from previous questions in the website. With the popularity of Stack Overflow, the founders of the website apply the same pattern to other fields, such as cooking and photography. Each CQA site is called Stack Exchange. Stack Exchange covers a wide range of topics.

Stack Exchange owns a large amount of Q&A data, and website operators regularly expose their data for the purpose of research. Based on the Q&A datasets, Correa and Sureka's [45] study deleted questions in the website to remind community members not to ask low-quality questions. Beyer and Pinzger [46] studied Stack Overflow tags and looked for the similar tags and merging them to avoid tag overflow. Meanwhile, the datasets from Stack Overflow have been used in expert finding and expert recommendation research. Faisal et al. [35] applied the g-index to expert ranking. Yang et al. [47] propose Topic Expertise Model (TEM) for expert finding. In order to verify our method, we adopt the dataset under the coffee topic of Stack Exchange website to simulate. The statistics of the dataset are shown in Table 2.

5. Results and Discussion

Stack Exchange possesses a reputation system, and each user owns a reputation score. Actually, the reputation system of Stack Exchange community has strict evaluation standards. Generally, if users want to improve their reputation score, they need to post valuable questions in their professional fields for a long time and provide high-quality answers or comments for other questions. The reputation system of Stack Exchange community helps the community to stimulate the potential of users and form a virtuous circle. Therefore, the reputation score of Stack Exchange community users has great reference value for evaluating expert users. Reputation score comes from the comprehensive performances of the user. We compare and analyze the results of the experts ranking with reputation scores of the users.

In the experiment, we choose two expert ranking techniques for comparison. The first one is Expertise Rank [29], which is an expert ranking method based on link analysis. The users are ranked in Expertise Rank according to the Q&A relationship among the users. The second one is LeaderRank [42], which is an improvement of Page Rank algorithm. In addition, LeaderRank is an adaptive parameter-free algorithm. We adopt the accuracy P to measure the difference of expert ranking algorithms. P is calculated as

$$P = \frac{r_{\text{list}} \cap e_{\text{list}}}{\text{num}_{\text{list}}}, \quad (10)$$

where r_{list} denotes the user reputation ranking list, e_{list} denotes the experts ranking list, and num_{list} presents the number of experts in the list. Adopting the reputation ranking list as a benchmark, we select the top 30, 50, and 100 users respectively from the experts ranking lists obtained by different methods to calculate the accuracies, the results are shown in Figure 4. Exp-Rank model combines the persistent performance and authority of users. The result shows that the expert list ranked by Exp-Rank has a high correlation with the user reputation ranking list. It indicates that expert users selected by Exp-Rank are generally recognized. In addition, the performances of both Expertise Rank and LeaderRank algorithms are not satisfied, principally because the evaluation indexes of these techniques are not comprehensive.

In the section of the expert recommendation, we obtain the top 100 expert users from the expert ranking list of Exprank and extract their knowledge tags. In the Stack Exchange dataset, more than 90% of the questions possess five answers at most, and the high-score answers are rare. Therefore, we select five questions with the most answers as the new questions, and these questions have their own tags. Though comparing the similarity between the new question tags and the knowledge tags from the top 100 expert users, the expert users can be further screened according to the similarity score. We compare the recommended expert users with the users who actually answered these questions and calculate the accuracy of the expert recommendation. The accuracy is calculated as

$$P = \frac{R_{\text{list}} \cap A_{\text{list}}}{R_{\text{list}}}, \quad (11)$$

$$\overline{P} = \frac{\sum_{i=1}^{N} P_i}{N}, \quad (12)$$

where R_{list} indicates recommended list of experts and A_{list} indicates the list of users who actually answered the

Table 1: User knowledge Tags descriptions.

User ID	User knowledge tags
1	Moka-pot aeropress latte-art bean-varieties egg-coffee
2	Diner-coffee caffeine cholesterol percolator soya-milk
3	French-press latte espresso-machine space kahlúa
4	Cafetiere Indian coffee-maker French-press acidic

Table 2: Stack Exchange dataset statistics.

Users	7510
Questions	1059
Answers	2107
Comments	4602

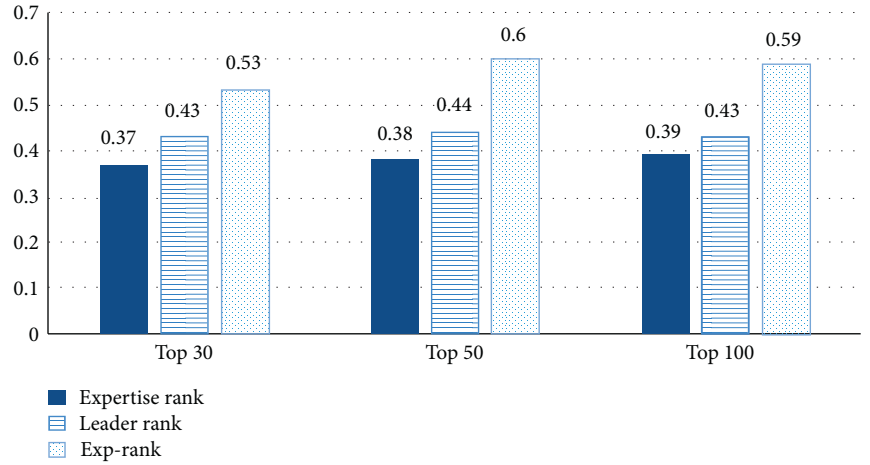


Figure 4: Accuracy of the top users.

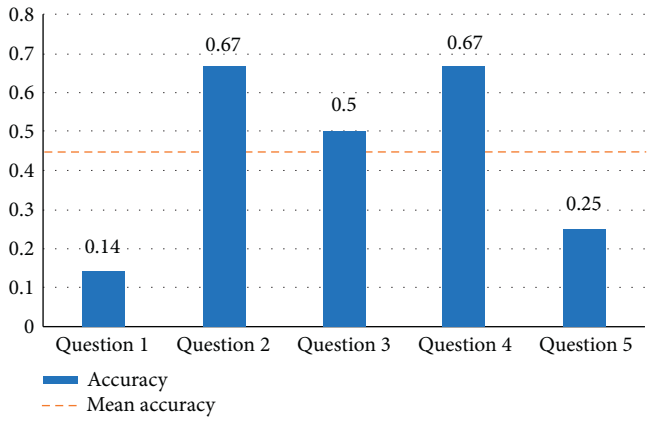


Figure 5: Accuracy score of recommended experts.

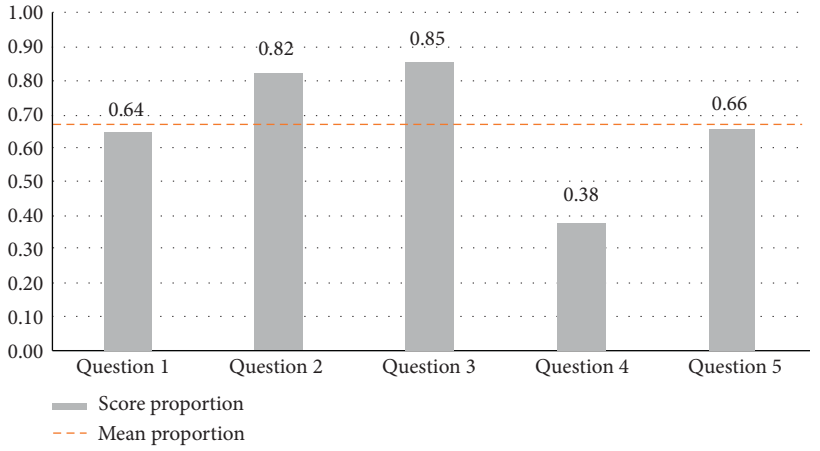


Figure 6: Percentage of the score for the expert's answers.

question. $\overline{P}$ is the mean accuracy. As shown in Figure 5, we recommend experts for the top five questions. The average accuracy is 0.44. That means when we recommend 10 experts, at least 4 of them will answer questions with high probability. As shown in Figure 6, for the five questions, the average proportion of expert users' answers scores is 0.67, which indicates that the answers of expert users are recognized by other users and have a higher professional level.

In conclusion, the experimental results prove that a better expert ranking result will be obtained by combining the authority of users with the continuous performance of users. In addition, recommending experts according to the similarity of the new questions can improve the accuracy of expert finding system.

6. Conclusion and Future Work

In order to identify the expert users in the complex online community, we propose a novel expert finding system based on the characteristics of the CQA sites. In our scheme, we propose an expert ranking algorithm named Exp-Rank, which considers the continuous performance and the authority of users and gives a more objective and comprehensive ranking of experts. Furthermore, we recommend experts according to the similarity between the new question and the knowledge tags of expert users. In particular, we can obtain some better results when we recommend 10–20 users. It should be noted that the evaluation indexes adopted in our method are common in CQA sites, so it can be widely applied in different types of online communities, such as Yahoo Answers and Zhihu. Consequently, we will try to enhance the performance of schemes with some more complex factors, including the user activity and the cold start problems of new users.

Data Availability

Previously reported Stack Exchange dataset was used to support this study and is available at https://archive.org/download/stackexchange. The dataset is freely available for research and has been used for finding experts and quality answers.

Conflicts of Interest

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

Acknowledgments

This work was supported by the Key Industry Projects in Shaanxi Province (no. 2019ZDLGY09-03) and the 111 Project under (Grant no. B08038).

References

[1] J. Bian, Y. D. Liu, D. Zhou et al., "Learning to recognize reliable users and content in social media with coupled mutual reinforcement," in *Proceedings of the Web Conference*, pp. 51–60, Madrid, Spain, April 2009.

[2] M. Bouguessa, B. Dumoulin, S. R. Wang et al., "Identifying authoritative actors in question-answering forums: The case of Yahoo! answers," in *Proceedings of the Knowledge Discovery and Data Mining*, pp. 866–874, Las Vegas, NV, USA, August 2008.

[3] A. Kardan, M. Garakani, and B. Bahrani, "A method to automatically construct a user knowledge model in a forum environment," in *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 717-718, Geneva, Switzerland, July 2010.

[4] F. Liu, S. Xue, J. Wu et al., "Deep learning for community detection: progress, challenges and opportunities," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence*, pp. 4981–4987, Yokohama, Japan, July 2020.

[5] A. Barua, S. W. Thomas, and A. E. Hassan, "What are developers talking about? An analysis of topics and trends in stack overflow," *Empirical Software Engineering*, vol. 19, no. 3, pp. 619–654, 2014.

[6] F. Calefato, F. Lanubile, and N. Novielli, "How to ask for technical help? Evidence-based guidelines for writing questions on stack overflow," *Information and Software Technology*, vol. 94, pp. 186–207, 2018.

[7] L. Adamic, J. Zhang, E. Bakshy et al., "Knowledge sharing and Yahoo answers: Everyone knows something," in *Proceedings of the Web Conference*, pp. 665–674, Beijing, China, April 2008.

[8] H. Lin, W. Fan, L. Wallace et al., "An empirical study of web-based knowledge community success," in *Proceedings of the Hawaii International Conference on System Sciences*, p. 178, Big Island, HI, USA, January 2007.

[9] F. Xiong, W. Shen, H. Chen, S. Pan, X. Wang, and Z. Yan, "Exploiting implicit influence from information propagation for social recommendation," *IEEE Transactions on Cybernetics*, vol. 50, no. 10, pp. 4186–4199, 2020.

[10] Z. Li, F. Xiong, X. Wang, H. Chen, and X. Xiong, "Topological influence-aware recommendation on social networks," *Complexity*, vol. 2019, pp. 1–12, Article ID 6325654, 2019.

[11] F. Xiong, X. Wang, S. Pan, H. Yang, H. Wang, and C. Zhang, "Social recommendation with evolutionary opinion dynamics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, pp. 1–13, 2018.

[12] J. Wu, X. Zhu, C. Zhang, and P. S. Yu, "Bag constrained structure pattern mining for multi-graph classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2382–2396, 2014.

[13] J. Wu, S. Pan, X. Zhu et al., "Boosting for multi-graph classification," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 416–429, 2015.

[14] B. Yang and S. Manandhar, "Exploring user expertise and descriptive ability in community question answering," in *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pp. 320–327, Beijing, China, August 2014.

[15] S. Lin, W. Hong, D. Wang, and T. Li, "A survey on expert finding techniques," *Journal of Intelligent Information Systems*, vol. 49, no. 2, pp. 255–279, 2017.

[16] A. M. Zuhair, K. Seifedine, and O. A. Isiaka, "Understanding expert finding systems: Domains and techniques," *Social Network Analysis & Mining*, vol. 8, no. 1, pp. 1–9, 2018.

[17] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, 1998.

[18] R. Divya, C. Aswani Kumar, S. Saijanani et al., "Deceiving communication links on an organization email corpus,"

Malaysian Journal of Computer Science, vol. 24, no. 1, pp. 17–33, 2011.

[19] D. M. Blei, A. Y. Ng, M. I. Jordan et al., "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[20] A. Tehmina, D. Ying, D. Ali et al., "Topic-based heterogeneous rank," *Scientometrics*, vol. 104, no. 1, pp. 313–334, 2015.

[21] A. El-korany, "Integrated expert recommendation model for online communities," *International Journal of Web & Semantic Technology*, vol. 4, no. 4, pp. 19–29, 2013.

[22] J. Yang, A. Bozzon, G. J. Houben et al., "E-WISE: An expertise-driven recommendation platform for web question answering systems," in *Proceedings of the International Conference on Web Engineering*, pp. 691–694, Rotterdam, The Netherlands, June 2015.

[23] A. Bozzon, M. Brambilla, S. Ceri et al., "Choosing the right crowd: expert finding in social networks categories and subject descriptors," in *Proceedings of the 16th International Conference on Extending Database Technology*, pp. 637–348, Genoa, Italy, March 2013.

[24] Z. Jing, T. Jie, and J. Li, "Expert finding in a social network," in *Proceedings of the Database Systems for Advanced Applications*, pp. 1066–1069, Bangkok, Thailand, April 2007.

[25] F. Yi, S. Luo, and A. P. Mathur, "Discriminative probabilistic models for expert search in heterogeneous information sources," *Information Retrieval*, vol. 14, no. 2, pp. 158–177, 2011.

[26] Y. P. Fu, R. J. Xiang, Y. Q. Liu et al., "Finding experts using social network analysis," in *Proceedings of the Web Intelligence*, pp. 77–80, Silicon Valley, CA, USA, November 2007.

[27] K. Balog and R. M. De, "Determining expert profiles (with an application to expert finding)," in *Proceedings of the International Joint Conference on Artifical Intelligence*, pp. 2657–2662, Hyderabad, India, January 2007.

[28] P. Jurczyk and E. Agichtein, "HITS on question answer portals: An exploration of link analysis for author ranking," in *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 845-846, Amsterdam The Netherlands, July 2007.

[29] J. Zhang, M. Ackerman, and L. Adamic, "Expertise networks in online communities: Structure and algorithms," in *Proceedings of the International Conference on World Wide Web*, pp. 221–230, Alberta, Canada, May 2007.

[30] R. R. Yang and J. H. Wu, "Study on finding experts in community question-answering system," *Applied Mechanics and Materials*, vol. 513–517, pp. 1760–1764, 2014.

[31] Z. Zhao, L. Zhang, X. He, and W. Ng, "Expert finding for question answering via graph regularized matrix completion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 4, pp. 993–1004, 2015.

[32] Ç. Aslay, N. O'Hare, L. Aiello et al., "Competition-based networks for expert finding," in *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1033–1036, Dublin Ireland, August 2013.

[33] B. Shao and J. Yan, "Recommending answerers for stack overflow with LDA model," in *Proceedings of the Conference on Computer Supported Cooperative Work*, pp. 80–86, Portland, OR, USA, March 2017.

[34] Y. Lu, X. Quan, J. Lei et al., "Semantic link analysis for finding answer experts," *Journal of Information Science and Engineering*, vol. 28, no. 1, pp. 51–65, 2012.

[35] M. S. Faisal, A. Daud, and A. Akram, "Expert ranking using reputation and answer quality of co-existing users," *International Arab Journal of Information Technology*, vol. 14, no. 2, pp. 118–126, 2017.

[36] A. U. Akram, K. Iqbal, C. M. S. Faisal et al., "An effective experts mining technique in online discussion forums," in *Proceedings of the International Conference Computing Electronic and Electrical Engineering*, pp. 95–99, Piura, Peru, 2016.

[37] F. M. Shahzad, D. Ali, A. A. Usman et al., "Expert ranking techniques for online rated forums," *Computers in Human Behavior*, vol. 100, pp. 168–176, 2018.

[38] W. C. Kao, D. R. Liu, and S. W. Wang, "Expert finding in question-answering websites: a novel hybrid approach," in *Proceedings of the ACM Symposium on Applied Computing*, pp. 867–871, Sierre, Switzerland, March 2010.

[39] G. A. Wang, J. Jiao, A. S. Abrahams, W. Fan, and Z. Zhang, "ExpertRank: A topic-aware expert finding algorithm for online knowledge communities," *Decision Support Systems*, vol. 54, no. 3, pp. 1442–1451, 2013.

[40] M. Rafiei and A. A. Kardan, "A novel method for expert finding in online communities based on concept map and PageRank," *Human-centric Computing and Information Sciences*, vol. 5, no. 1, pp. 1–18, 2015.

[41] G. Zhou, S. Lai, K. Liu et al., "Topic-sensitive probabilistic model for expert finding in question answer communities," in *Proceedings of the Conference on Information and Knowledge Management*, pp. 1662–1666, Maui, HI USA, October 2012.

[42] L. Lü, Y. Zhang, C. H. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," *PLoS One*, vol. 6, no. 6, Article ID e21202, 2011.

[43] S. J. Rose, D. Engel, N. Cramer et al., "Automatic keyword extraction from individual documents," *Text Mining: Applications and Theory*, pp. 1–20, 2010.

[44] A. Furtado, N. Oliveira, and N. Andrade, "A case study of contributor behavior in Q&A site and tags: The importance of prominent profiles in community productivity," *Journal of the Brazilian Computer Society*, vol. 20, no. 16, p. 5, 2014.

[45] D. Correa and A. Sureka, "Chaff from the wheat: characterization and modeling of deleted questions on stack overflow," in *Proceedings of the Web Conference*, pp. 631–642, Toulouse, France, July 2014.

[46] S. Beyer and M. Pinzger, "Grouping android tag synonyms on stack overflow," *Mining Software Repositories*, vol. 11, pp. 430–440, 2016.

[47] L. Yang, M. Qiu, S. Gottipati et al., "CQARank: Jointly model topics and expertise in community question answering," in *Proceedings of the Conference on Information and Knowledge Management*, pp. 99–108, San Francisco, CA. USA, October 2013.