https://doi.org/10.31449/inf.v48i19.6544 Informatica **48** (2024) 73–88

Breadth Learning and Network Embedding Techniques for Sensitive Information Detection in Social Media

Jianping Luo[1*], Rongshen Zhu[2]
[1]School of Arts and Science, Chengdu College of University of Electronic Science and Technology of China, Chengdu 611731, China
[2]School of Computer Science, Chengdu College of University of Electronic Science and Technology of China, Chengdu 611731, China
E-mail: luojianpingcn@126.com
[*]Corresponding author

Keywords: sensitive information, network security, breadth learning, heterogeneous social networks, network embedding

Received: July 3, 2024

Social media platforms have brought tremendous changes to the way society communicates and transmits information. However, some extreme individuals post sensitive information such as false, malicious attacks, terrorism, or negative comments. To address this rapidly evolving network security challenge, this study introduces breadth learning and network embedding techniques from machine learning to establish a new method for information extraction and auditing. This method adopts a meta path random walk form to extract and filter effective information from multiple heterogeneous social networks. A network embedding technique based on breadth learning is proposed to transform each node into a low-dimensional feature space. Finally, a fusion function is designed to quantify the correlation between sensitive users on social platforms. The relationship between users and sensitive items is predicted by the correlation score, and then the sensitive information publisher is screened. Finally, the sensitive information released by the sensitive information publisher is effectively locked after it is identified. As a result, a sensitive information extraction and review model based on breadth learning and network embedding is constructed. The results showed that the research model exhibited a high degree of accuracy (97.3%) in mixed datasets on social platforms, outperforming other models designed for the review of sensitive information. Moreover, the model demonstrated superior performance on social platforms with larger data volumes. Accuracy represents the percentage of correct identification of sensitive information, indicating that the research model can more accurately identify sensitive information and its publishers. The research results have significant implications for the network security governance of open large-scale social media platforms.

Povzetek: Prispevek predstavlja novo metodo za zaznavanje občutljivih informacij na družbenih omrežjih, ki temelji na tehniki strojnega učenja in vdelavi omrežja. Metoda izboljša pregled občutljivih informacij, kar prispeva k varnosti omrežij.

1 Introduction

Social media on the Internet has brought convenience to interpersonal communication. However, the release review mechanism of these platforms is not perfect, causing a large number of sensitive information released by extreme people in the diversified information of social media [1]. Based on K-nearest neighbor algorithms or improved convolutional neural networks in machine learning, sensitive information, and information publishers can be accurately identified from the network. However, these methods are unable to collaboratively process and analyze sensitive information data across multiple social media platforms [2-3]. Breadth Learning (BL) technology in deep learning can integrate data from multiple different sources and mine the required information from them. It does not require deep structure and gradient descent to update weights, its calculation speed is extremely fast and can be used for real-time detection of sensitive information from multiple platforms [4]. Network Embedding Technology (NET) uses vector space to represent network information, which can reduce the dimensionality of network big data. It can randomly walk in the network, capture the relationships between information in the network topology, link multiple social media, and achieve network alignment of Sensitive Information Publishers (SIPs) on multiple platforms [5-7]. Breadth Learning and Network Embedding (BL-NE) technology is widely used in social media analysis and network recommendations. Many scholars have conducted relevant research on BL, NET, and social media Sensitive Information Review (SMSIR).

Zhang et al. proposed an Air Quality Index (AQI) prediction model that combines a decomposition algorithm and generalized learning to address the issue of inaccurate air pollution index prediction. This model outperformed other decomposition algorithms and improved prediction accuracy by 16.9%. It has been proven that integrating multiple air pollution data sources with BL could provide more accurate AQI predictions [8]. To solve the problem of low accuracy in image classification caused by the high complexity of hyperspectral images, Xiao et al. designed a hyperspectral image classification method based on Hierarchical Breadth Learning (HBL). HBL outperformed other methods in terms of classification accuracy and ability to recognize complex samples [9]. In response to the difficulty in finding effective attack points from Heterogeneous Combat Networks (HCNs) in distributed combat systems, Zeng et al. constructed an HCN attack sequence search model that combines NET and SHATTER. The performance of the proposed search model has improved by 31% compared to other technologies, and NET has solved the problem of searching for attack targets in HCN [10]. It was necessary to address the issue of maintaining the connectivity of embedded networks to ensure that cloud-based substrate networks can provide services to various virtual network clients. Therefore, Ogino proposed an elastic virtual NET that can ensure the connectivity of embedded virtual networks after critical node failures. This study conducted simulation experiments and verified that the technique exported VNE solutions with an approximate ratio of less than 1.2 within 10 s, indicating that NET can solve problems related to multi-platform networks [11]. Shen et al. proposed a communication network system based on satellite collaborative work to address the issue of the inability to locate SIPs and the inability to quickly connect cyberspace and the real world in cybersecurity. The system achieved the integration of time and space, creating a mathematical and physical endogenous security environment, which brings significant breakthroughs to network security issues [12]. The use of semantic segmentation for sensitive information recognition was prone to the problem of missing short information, and the real-time performance of deep networks was poor. Given this, Zhang and Ma established an improved semantic segmentation method for identifying sensitive information using Cloth Net-tiny based on the Atrus spatial pyramid pool. The average intersection of this method's network was 78.8% higher than the joint mean, and the memory consumption of the fully convolutional network was reduced by 77%. This indicated that the improved method improved the accuracy of detecting short information and consumed less computational resources [13]. The summary table of the reviewed research is shown in Table 1.

Table 1: The summary table of the reviewed research

Reference	Key findings	Methodologies	Performance metrics
Zhan et al [8]	The accuracy of AQI prediction has increased by 16.9%	AQI Prediction Model Combining Decomposition Algorithm and Generalized Learning	Integrate multiple air pollution data sources to improve prediction accuracy
Xiao et al [9]	The accuracy of hyperspectral image classification has been improved	Hyperspectral Image Classification Method Based on HBL	Using BLS and Gabor filtering alternately can explore the spatial information of images
Zeng et al [10]	Performance improvement of 31% in finding effective attack points in HCN of distributed combat systems	An HCN attack sequence search model combining NET and SHATTER	Combining inductive algorithms, GNN, and reinforcement learning to solve the problem of network disintegration sequences
Ogino [11]	The research technology has exported a VNE solution with an approximate ratio less than 1.2 within 10 seconds, which can maintain the connectivity of embedded networks	Propose a resilient virtual NET	Research technology can ensure the connectivity of embedded virtual networks after critical node failures occur
Shen et al [12]	Create an endogenous security environment for mathematics and physics, bringing significant breakthroughs to network security issues	A communication network system based on satellite collaborative work has achieved the integration of time and space	Integrating the Internet and GPS, realizing the integration of physical time and space
Zhang and Ma [13]	The average intersection of the recognition method's network layout is 78.8% higher than the joint average, and the memory consumption is reduced by 77% compared to fully convolutional networks	Improved semantic segmentation method for Cloth Net tiny recognition of sensitive information based on atrous spatial pyramid pool	The improved semantic segmentation method enhances the accuracy of detecting short information and consumes less computational resources

The above research indicates that many scholars have conducted relevant studies on BL, NET, and SMSIR. There is little research on using BL-NE technology for social media information extraction and review. Therefore, this study combines BL-NE technology to construct a new SMSIR method. In this study, social media platforms will be heterogeneous, and Metapath Random Walk (MPRW) will be used to extract and filter effective information. Based on effective information, a BL-based NET is used to embed and fuse sensitive nodes, and finally, the correlation score is used to identify the SIPs. The innovation of the research lies in the use of multi-platform heterogeneous NET fusion matrix factorization, which quantifies the correlation score to achieve accurate identification of SIP.

The research has five parts. Part 1 is the introduction, which analyzes the research achievements in social media information review and briefly describes the proposed SMSIR model. Part 2 is an algorithm and mechanism for SMSIR and information extraction based on BL-NE technology. Part 3 tests the research model. Part 4 discusses the experimental results. Part 5 summarizes the research findings.

2 Methods and materials

To accurately link the relationships between network users, this study introduces Heterogeneous Social Networks (HSNs) to represent users, platforms, and information releases on social platforms in the form of nodes. Metapath is used to abstractly represent the link relationships between network nodes [14-15]. This study first points out the network alignment methods in HSNs, and then designs methods to extract and filter sensitive information from HSNs. Afterwards, a heterogeneous social NET based on BL is proposed, which identifies sensitive information and its publishers through correlation scores. Finally, the parameter optimization of the recognition model is completed, and an SMSIR model based on BL-NE is constructed.

2.1 Methods for extracting and filtering effective information in social media

Social media platforms come in various forms with low registration requirements. SIPs usually register accounts on multiple platforms and publish sensitive information on one platform. Numerous social participants form a complex network of relationships on social media. HSNs can accurately and efficiently represent different users and their relationships in a network. Metapath can intuitively and efficiently abstract the types and quantities of users and links in social networks. The schematic diagram of HSNs is shown in Fig. 1.

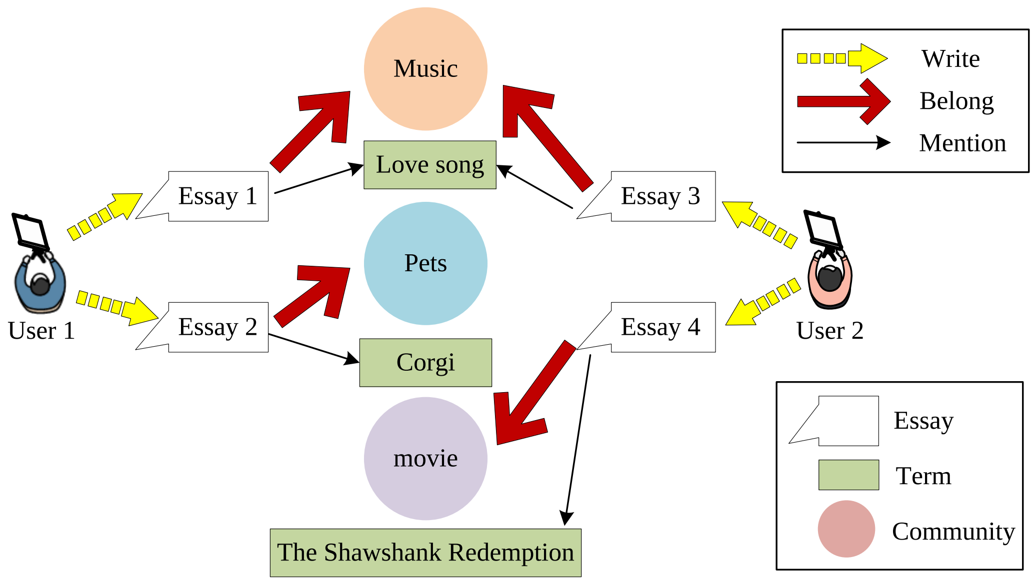


Figure 1: Schematic diagram of heterogeneous social networks

In Fig. 1, HSNs contain nodes of social platforms, users, information, and keywords. There is a relationship between nodes where information belongs to social platforms, users write information, and information has keywords. In HSNs, different users will post information related to platform interests on different social platforms. By mining information from HSNs, accurate and comprehensive user relationships and information transmission on social media can be obtained. Therefore, this study proposes a strategy for SIP identification from multi-platform heterogeneous networks, which can more accurately identify SIP. Fig. 2 is a Metapath diagram of obtaining network pattern instances from HSNs.

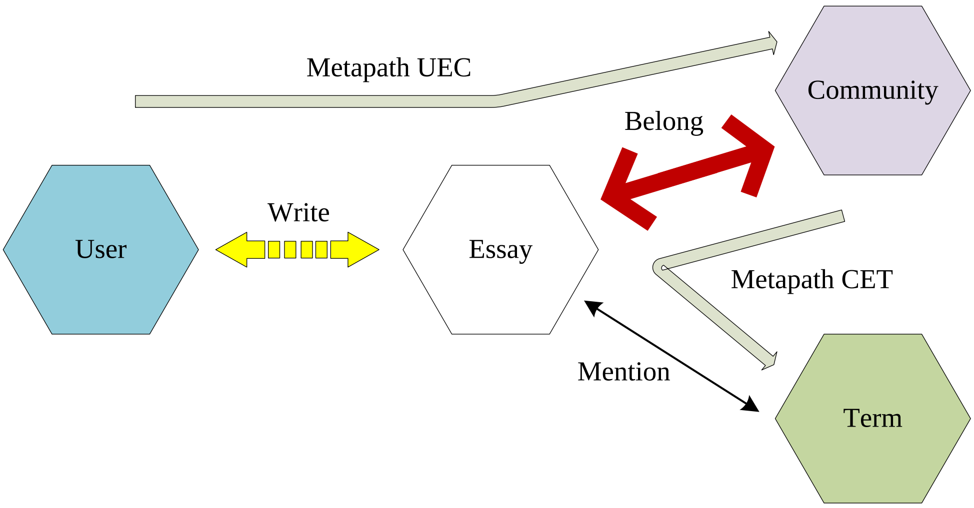


Figure 2: Schematic diagram of obtaining network pattern instance meta paths from HSNs

In Fig. 2, a Metapath is represented in the order of multiple nodes, abbreviated as a Metapath. CET represents the path from information to keywords on social platforms. UEC represents the path where the publisher writes information and posts it on social media platforms. Metapath can effectively express the connection between information and users on social platforms. The process corresponding to SIP on different social platforms is called network alignment [16-17]. At present, the latest research on network security mainly focuses on the network alignment of two HSNs. The general expression for network alignment is equation (1).

$$f:\left(G^1,G^2\right)\rightarrow \left\{\left(V^1,V^2\right)\middle| V^1\in G^1,V^2\in G^2,V^1=V^2\right\} \tag{1}$$

In equation (1), V^n and E^n are the sets of nodes and Metapaths in HSNs, respectively, and $n=\{1,2\}$. R^n is the $|V^n|\times d^n$ -dimensional attribute matrix in HSNs, and $G^n=\left(V^n,E^n,R^n\right)$. The specific network alignment forms mainly adopt three types: topic similarity, username, and user attributes. Among them, username similarity is the fastest and simplest method for user network alignment. Information publishers usually use the same or similar username to register on multiple social media platforms, and the calculation formula for similarity judgment using the username is equation (2).

$$sim_{name}(u,v)=\sum_{i=1}^{n}\left(NP_i\times\omega_i\right) \tag{2}$$

In equation (2), u and v represent the usernames registered by the user on different platforms. NP_i is the set of similarities between the attributes of two usernames. ω_i represents the weight of the i -th attribute, $\omega_i=(0,1)$, and the sum of all weights is 1. The larger the $sim_{name}(u,v)$, the greater the probability that two platform users are the same user. On social media platforms, users usually fill in attribute information such as birthday, location, interests, etc. The calculation formula for determining user similarity using these attributes is equation (3).

$$sim_{background}(u,v)=\sum_{i=1}^{n}\left(BP_i\times\omega_i\right) \tag{3}$$

In equation (3), BP_i represents the similarity set of various background attributes between users u and v. The larger the $sim_{background}(u,v)$, the greater the probability that two platform users are the same user. In addition, there may be false information in the background attributes, which can reduce the accuracy of this method's judgment. The information published by the same user on different platforms usually has a certain degree of similarity, and by mining the information, user alignment based on similar interest topics can be achieved. The calculation formula for determining user similarity using interest topics is equation (4).

$$sim_{inerest}(u,v)=\sum_{i=1}^{n}getsim\left(userl_i word_i,user2_i word_i\right) \tag{4}$$

In equation (4),

$$Userl=\left\{iword_1,iword_2,...,iword_n\right\}$$

represents the set of keywords in the information published by users u and v among various HSNs. The larger the $sim_{inerest}(u,v)$, the greater the probability that two users are the same user. In social platforms, a considerable proportion of users access the platform for information only, without providing any identity attributes or posting messages. This results in the aforementioned methods being limited by the lack of available information. However, users may leave likes and follow behaviors. By mining the interest labels behind behavior, the network structure of users can be constructed. Using the similarity of user network structure can achieve network alignment for different users. The formula for calculating user interest classification is equation (5).

$$
\begin{aligned}
&UserITword_n = \\
&\{itword_1, itword_2, \cdots, itword_n\} \Rightarrow \\
&UserITC_n = \{ITC_1, ITC_2, \cdots, ITC_n\}
\end{aligned}
\tag{5}
$$

In equation (5), it_n is a set of interests. $ITC = \{it_1, it_2, \cdots, it_n\}$ represents the classification set of user interests mined. $itword_n$ is the user's interest word. The expression for determining user similarity based on user interests is equation (6).

$$
sim_{structure}(u,v) = \frac{UserITC_1 \cup UserITC_2}{UserITC_1 \cap UserITC_2} \tag{6}
$$

In equation (6), the higher the $sim_{structure}(u,v)$, the greater the probability that two users are the same user. The method of using network structure for network alignment is more accurate when there is less effective SIP information. To achieve network alignment and recognition of sensitive information, this study uses the MPRW method to extract effective information from HSNs. The formula for the random walk path is equation (7).

$$
P\left(n_{t+1} = x \mid n_t = v, \rho\right) =
\begin{cases}
\dfrac{1}{\left|N^{A_{t+1}}(v)\right|} & (v,x) \in \mathrm{E} \; and \; \phi(x) = A_{t+1} \\
0, & otherwise
\end{cases}
\tag{7}
$$

In equation (7), $\phi_k \in \{\phi_0, \phi_1, \phi_2, ..., \phi_n\}$ represents the Metapath of HSNs. $N^{A_{t+1}}(v)$ is the set of all adjacent nodes of user v. A_{t+1} is the type of user. n_t is the t-th node in a random walk. Fig. 3 is a schematic diagram of generating and extracting effective information through random walks in HSNs.

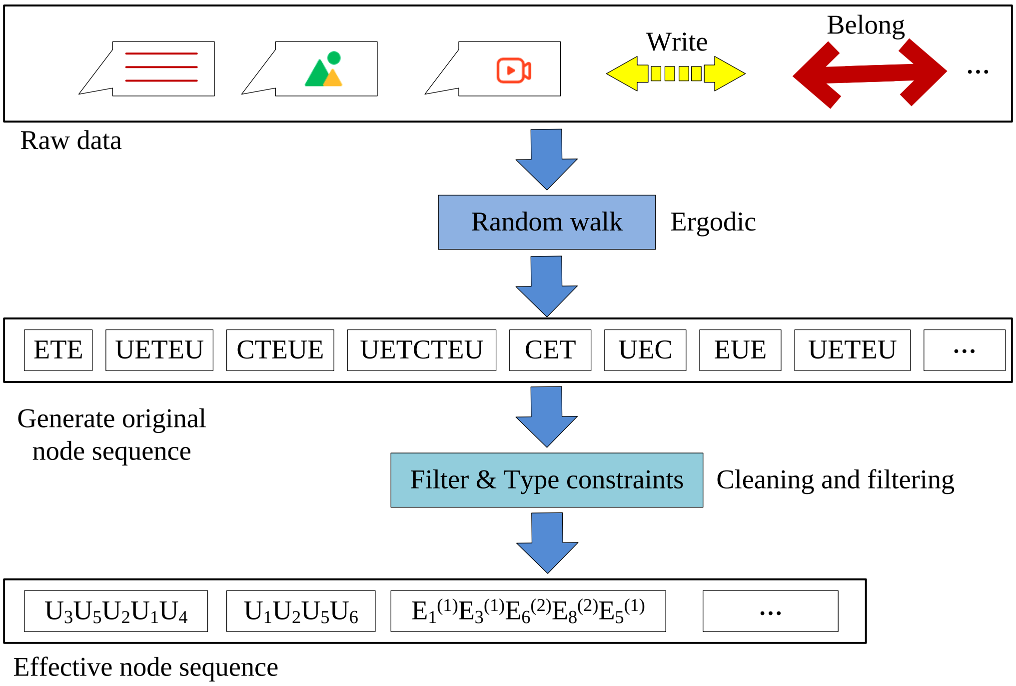


Figure 3: A random walk to generate and extract a diagram of valid information

In Fig. 3, the original network structure data include nodes such as text, images, and videos, as well as the relationships between nodes. A sequence of the original nodes is generated through random walks. This study adopts the method of cleaning and filtering node sequence data for data dimensionality reduction, so that the reduced data only contains sensitive items and users, and nodes of the same type are in the same sequence.

2.2 Construction of sensitive information audit model based on BL-NE

Traditional machine learning algorithms are difficult to handle large-scale, highly complex structures, and diverse attributes in multi-platform HSNs. Based on BL, NET can be embedded concurrently across multiple social platforms, extracting information from mature social networks for information exchange and symmetry with new platforms. This technology can solve the problem of information sparsity and construct richer features of SIP by refining and embedding between platforms. BL-based NET first defines a set of element proximity for Metapaths in HSNs. The formula for defining the user's meta proximity is equation (8).

$$\gamma_{\phi_k}^{(n)}(u_1^{(n)},u_2^{(n)}) = \frac{2\,|\,P_{\phi_k}^{(n)}(u_1^{(n)},u_2^{(n)})\,|}{|\,P_{\phi_k}^{(n)}(u_1^{(n)},\cdot)\,|+|\,P_{\phi_k}^{(n)}(\cdot,u_2^{(n)})\,|} \quad (8)$$

In equation (8), $\gamma_{\phi_k}^{(n)}(u_1^{(n)},u_2^{(n)})$ represents the meta proximity between users $u_i^{(n)}$ and $u_j^{(n)}$. $P_{\phi_k}^{(n)}(u_1^{(n)},u_2^{(n)})$ is the set of Metapaths between users. $P_{\phi_k}^{(n)}(u_1^{(n)},\cdot)$ and $P_{\phi_k}^{(n)}(\cdot,u_2^{(n)})$ are the sets of all related Metapaths for $u_i^{(n)}$ and $u_j^{(n)}$ in the network. The definition of the proximity of sensitive information elements is equation (9).

$$\gamma_{\Phi_k}^{(n)}(i_1^{(n)},i_2^{(n)}) = \frac{2\,|\,P_{\Phi_k}^{(n)}(i_1^{(n)},i_2^{(n)})\,|}{|\,P_{\Phi_k}^{(n)}(i_1^{(n)},\cdot)\,|+|\,P_{\Phi_k}^{(n)}(\cdot,i_2^{(n)})\,|} \quad (9)$$

In equation (9), $\gamma_{\Phi_k}^{(n)}(i_1^{(n)},i_2^{(n)})$ is the meta proximity of the sensitive node. $P_{\Phi_k}^{(n)}(i_1^{(n)},\cdot)$ and $P_{\Phi_k}^{(n)}(\cdot,i_2^{(n)})$ are the sets of all Metapaths related to sensitive nodes. $P_{\Phi_k}^{(n)}(i_1^{(n)},i_2^{(n)})$ is the set of Metapaths between sensitive nodes. The $\gamma_{\phi_k}^{(n)}(u_1^{(n)},u_2^{(n)})$, $\gamma_{\Phi_k}^{(n)}(i_1^{(n)},i_2^{(n)})$, and other potentially sensitive information in HSNs are combined to learn all sensitive information in HSNs and feature vectors embedded in user networks. The $S_{\Phi_k}^{(1)}$ and $S_{\Phi_k}^{(2)}$ matrices represent the generated results. The i-th row in $S_{\Phi_k}^{(1)}$ and $S_{\Phi_k}^{(2)}$ represents the embedding feature vector of $u_i^{(n)}$ in HSNs. When $S_{\phi_k}^{(1)}(i,\cdot)$ and $S_{\phi_k}^{(2)}(\cdot,j)$ are mapped to close positions in a low dimensional space, it indicates that $u_i^{(n)}$ and $u_j^{(n)}$ are accounts registered by the same publisher on different HSNs. By embedding the results, a domain N_u is set for the user u in a low-dimensional space, and the embedding function is optimized using stochastic gradient descent. The embedded function expression is equation (10).

$$\max_f \sum \log \Pr\left(N_u \mid f(u)\right) \rightarrow f(\cdot) \quad (10)$$

In equation (10), $f(\cdot)$ is the embedding function. The embedding results of HSNs based on BL are shown in Fig. 4.

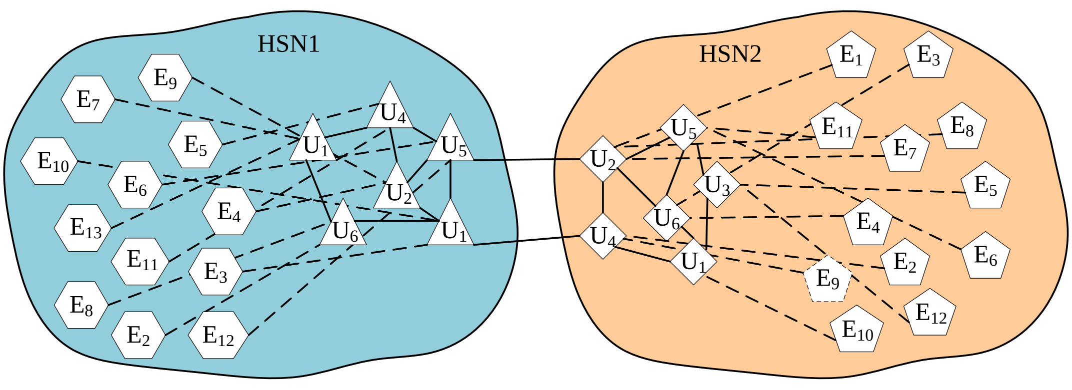


Figure 4: Embedding result graph of HSNs based on BL

In Fig. 4, HSN1 and HSN2 are two social platforms, and sensitive information and users are mapped to a low dimensional space through BL-based NET. When there is a situation where a SIP uses both HSN1 and HSN2, two similar accounts will be located in close proximity and have a linked relationship, as shown in figures U1-U4 and U5-U2. This study integrates the results of HSNs embedding into the matrix factorization framework to ultimately determine the SIP from the correlation score [18]. Firstly, a fusion function is used to learn the embedding vectors of users and sensitive information, and the final embedding expressions $\left\{e_i^{(I)}\right\}_{i\in I}$ and $\left\{e_u^{(U)}\right\}_{u\in U}$ of sensitive information and users are obtained. The calculation of user fusion function is equation (11).

$$g\left(\left\{e_u^{(U)}\right\}\right)=\sigma\left(\sum_{n=1}^{|P|}\omega^{(n)}\sigma\left(M^{(n)}e_u^{(U)}+b^{(n)}\right)\right) \tag{11}$$

In equation (11), σ is the Sifmoid function. $M^{(n)}$ is the transformation matrix. $|P|$ is the number of sets of Metapaths. $\omega^{(n)}$ is the correlation coefficient between user u and the n-th Metapath. The formula for the sensitive information fusion function is equation (12).

$$g\left(\left\{e_i^{(I)}\right\}\right)=\sigma\left(\sum_{n=1}^{|P|}\omega^{(n)}\sigma\left(M^{(n)}e_i^{(I)}+b^{(n)}\right)\right) \tag{12}$$

In equation (12), $b^{(n)}$ is the transformation matrix. According to the matrix factorization framework, the correlation score between sensitive information i and user u can be calculated. The calculation of correlation score is equation (13).

$$\hat{\Gamma}_{u,i}=x_u^T\cdot y_i \tag{13}$$

In equation (13), $\hat{\Gamma}_{u,i}$ represents the predicted score of the correlation between i and u. y_i and x_u^T are sensitive information and potential factors corresponding to users. The expression after introducing $\left\{e_i^{(I)}\right\}_{i\in I}$ and $\left\{e_u^{(U)}\right\}_{u\in U}$ into the correlation score calculation is equation (14).

$$\begin{aligned}&\hat{\Gamma}_{u,i}= \\ &x_u^T\cdot y_i+\alpha\cdot e_u^{(U)T}\cdot\lambda_i^{(I)}+\beta\cdot e_i^{(I)T}\cdot\lambda_u^{(U)}\end{aligned} \tag{14}$$

In equation (14), $\lambda_i^{(I)}$ and $\lambda_u^{(U)}$ are sensitive information and user specific potential factors, and α and β are fusion coordination parameters. Due to the setting of parameters has a significant impact on the accuracy of SIP recognition, a parameter optimization method is designed. This study uses the stochastic gradient descent method to optimize parameters, and the objective of parameter learning is equation (15).

$$\begin{aligned}&\kappa=\sum(\Gamma_{u,i}-\hat{\Gamma}_{u,i})^2+ \\ &\mu\sum(\|x_u\|_2+\|y_i\|_2+\|\lambda_u^{(U)}\|_2+\|\lambda_i^{(I)}\|_2+\|\Theta^{(U)}\|_2+\|\Theta^{(I)}\|_2)\end{aligned} \tag{15}$$

In equation (15), $\Gamma_{u,i}$ is the actual score of the correlation between i and u. $\Theta^{(I)}$ and $\Theta^{(U)}$ are parameters in the sensitive information and user fusion function. μ is the regularization parameter. The process of parameter optimization is shown in Fig. 5.

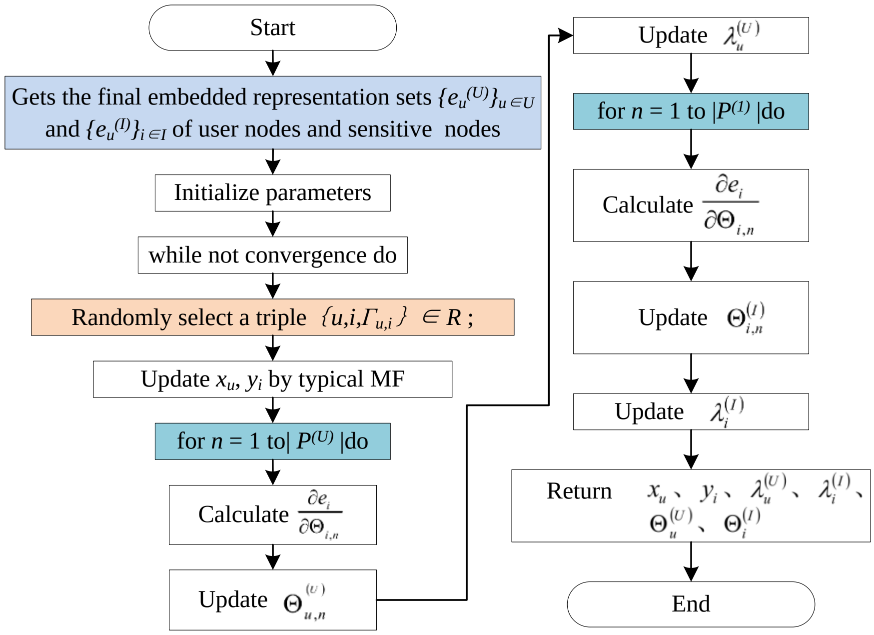


Figure 5: Flowchart for parameter optimization

In Fig. 5, the first is to input sensitive information and the final embedding set of the user to complete parameter initialization. Then, the random gradient descent optimization method is used to learn the parameters and complete the optimization. Finally, potential factors x_u and y_i for user and sensitive information are outputted, and specific potential factors $\lambda_i^{(I)}$ and $\lambda_u^{(U)}$ for pairing with $e_i^{(I)}$ and $e_u^{(U)}$ are outputted, as well as updated parameters $\Theta_u^{(U)}$ and $\Theta_i^{(I)}$. After parameter updates, a sensitive information and publisher recognition model that can be used for social media fusion network embedding and BL is obtained. The overall framework is shown in Fig. 6.

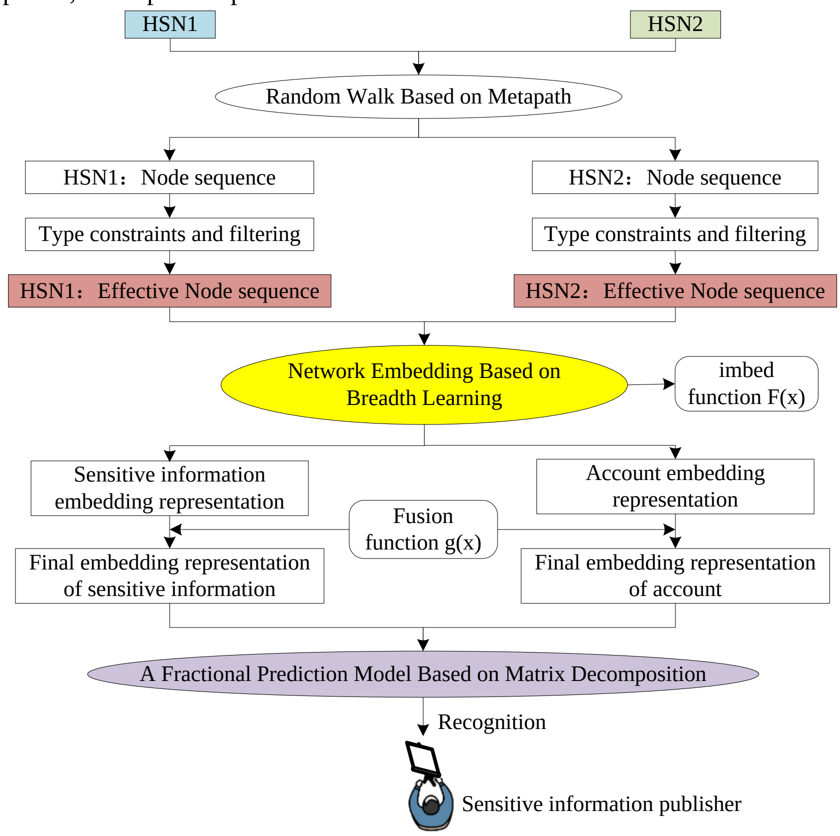


Figure 6: Sensitive information recognition model combining NET and BL

In Fig. 6, in the initial stage, the research model utilizes the MPRW mechanism to create an ordered node sequence and complete filtering, extracting effective information. Then, sensitive information and user nodes in the two HSNs networks are mapped in the same low--dimensional space through BL-NE. Finally, a fusion function is designed to learn the comprehensive embedding expression between users and sensitive items. It is integrated into a matrix decomposition model to predict the correlation score between the two, and then screen out hidden SIPs in the network. Auditing all sensitive information published by a locked publisher has extremely high accuracy.

3 Results

To verify the performance of the system, relevant experiments are conducted in this study. Due to the parameter settings of the model have a significant impact on the final review effect, the experiment first uses a single variable method to test the optimal value of the parameter settings. The optimal parameters for testing include fusion learning parameters α and β, the number of potential factors t, and the embedding dimension s. Finally, comparative experiments are conducted between the research model and other sensitive information review and recognition models to test the performance of the research model.

3.1 Experimental environment and parameter settings

The popular Instagram and Reddit social platforms have a large number of users, with a large daily active user base, high hidden SIP activity, and rich interactive information such as likes among users, making it easy to analyze user data. Therefore, this study collaborates with other teams to collect real social media data from Instagram and Reddit. The user-related information in the data includes username, gender, account, location, etc. Publish messages including content, topic, time, etc. Interactive behaviors include likes, comments, etc. Table 2 provides a detailed overview of social media data.

Table 2: Detailed information on parameter settings for algorithms

Instagram Database Table Name	Instagram field	Instagram Field type	Instagram Number of data	Reddit Database Table Name	Reddit field	Reddit Field type	Reddit Number of data
User 1	Uid 1	varchar (20)	3482	User 2	Uid 2	varchar(20)	6349
	Uname 2	varchar(50)			Uname 2	varchar(50)	
	Sex 1	varchar(10)			Sex 2	varchar(10)	
	Address 2	varchar(20)			Address 2	varchar(20)	
	Description2	varchar(100)			Description2	varchar(100)	
Item 1	Mid 2	varchar(20)	688	Item 2	Mid 2	varchar(20)	854
	Uid 2	varchar(20)			Uid 2	varchar(20)	
	Date 2	datetime			Date 2	datetime	
	Text 2	varchar(100)			Text 2	varchar(100)	
	Topic 2	varchar(50)			Topic 2	varchar(50)	
Interaction 1	Ctext 2	varchar(100)	25187	Interaction 2	Ctext 2	varchar(100)	59211
	Suid 2	varchar(20)			Suid 2	varchar(20)	
	Tmid 2	varchar(20)			Tmid 2	varchar(20)	
	Tuid 2	varchar(20)			Tuid 2	varchar(20)	
	Link 2	bit			Link 2	bit	
	Comment2	bit			Comment 2	bit	
	Reprint 2	bit			Reprint 2	bit	

The study randomly divides the above data into 12 subsets, with 8 as the training set and the remaining 4 as the testing set. The computer used in the experiment has 16GB of memory and an Intel i7 8700 processor. In the optimal parameter testing experiment, the range of potential factor t values is [0,30], the range of embedding dimension s values is [0256], and the range of fusion learning parameters α and β values is [0,2]. The experiment uses Hidden Markov Model (HMM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Bidirectional Encoder Representations from Transformers (BERT) to compare with the research model. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), accuracy, and F1 score are used as evaluation indicators for the accuracy of sensitive information recognition.

3.2 Optimal parameter analysis of sensitive information review model

The embedding dimension s is the most important parameter for studying model accuracy. This study tests the optimal parameter values of s on Instagram, Reddit, and mixed datasets. Taking a lower embedding dimension can reduce computational complexity, while taking a higher value can capture the complex structure and relationships between nodes in the network. Considering the complexity of the dataset, the research will expand the range as much as possible, with values ranging from 0 to 256. The experimental results are shown in Fig. 7.

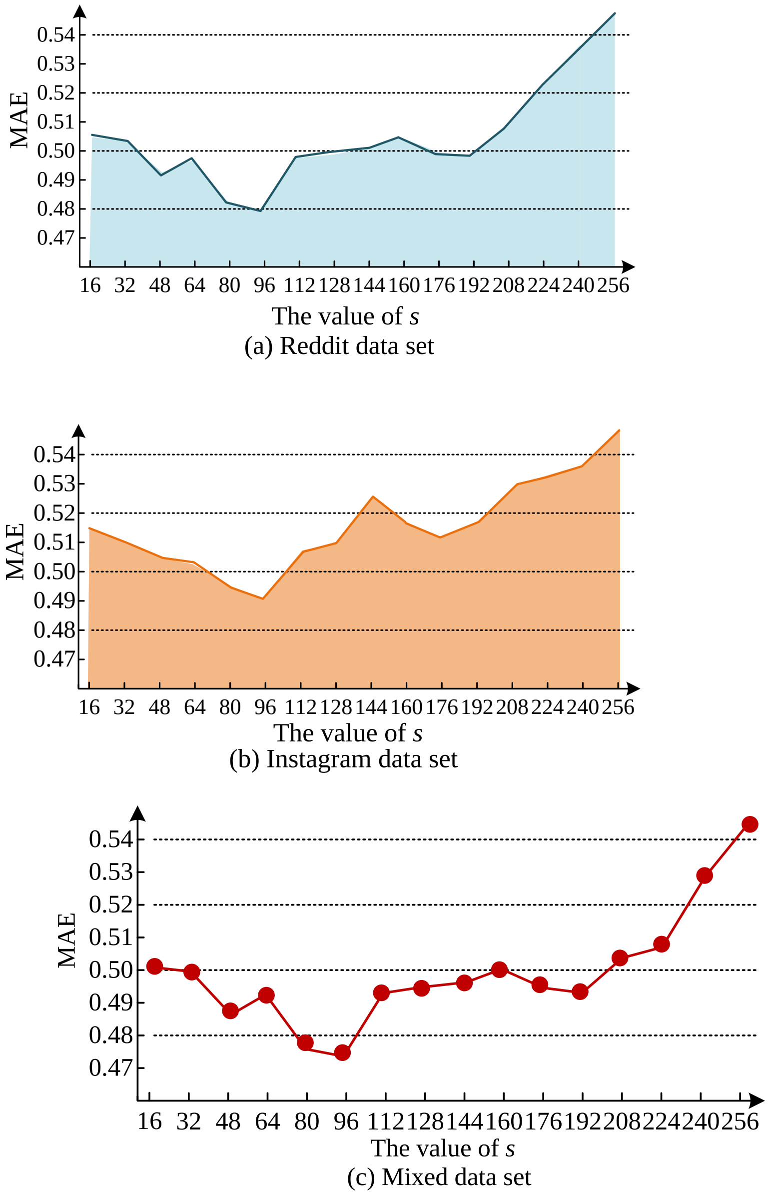
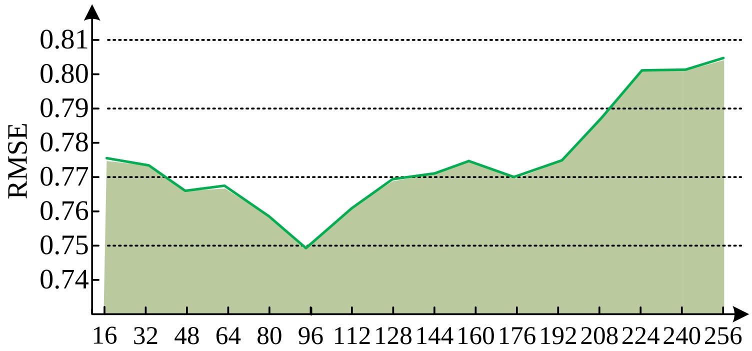
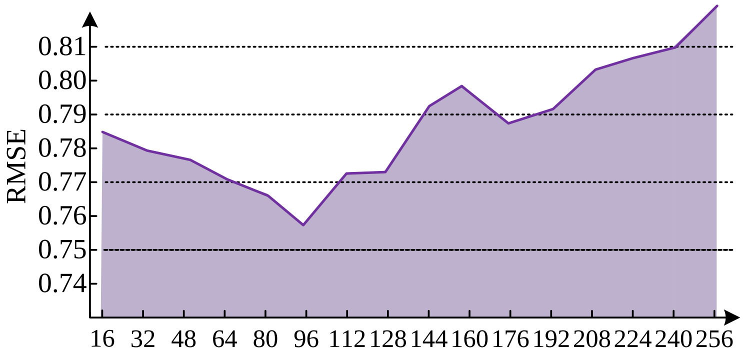


Figure 7: The influence of embedding dimension s on MAE

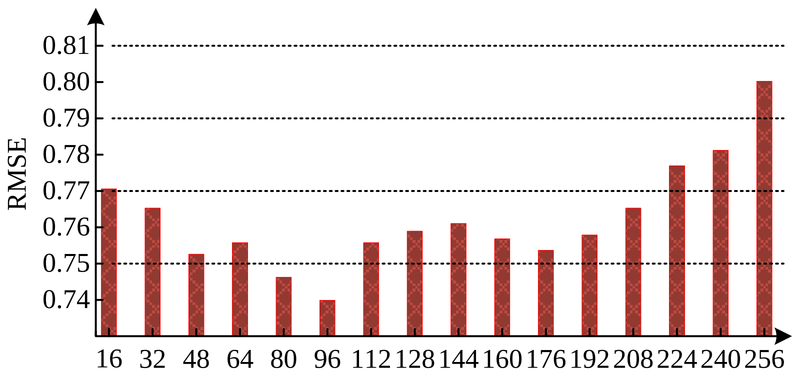
In the Reddit data set of Fig. 7 (a), both small and large embedding dimensions s will reduce the accuracy of the model, especially when the parameter value is 256 and the MAE value is 0.546. In Fig. 7 (b), the results of Instagram data testing are similar to those of Reddit data testing. In Fig. 7 (c), the value of s will cause fluctuations in MAE. In Figures 7 (a), (b), and (c), the minimum MAE values obtained at s=96 are 0.48, 0.492, and 0.474, respectively. To verify whether s=96 is the optimal parameter, the experiment continues to test the effect of the value of s on RMSE, as shown in Fig. 8.



(a) Reddit dataset test result



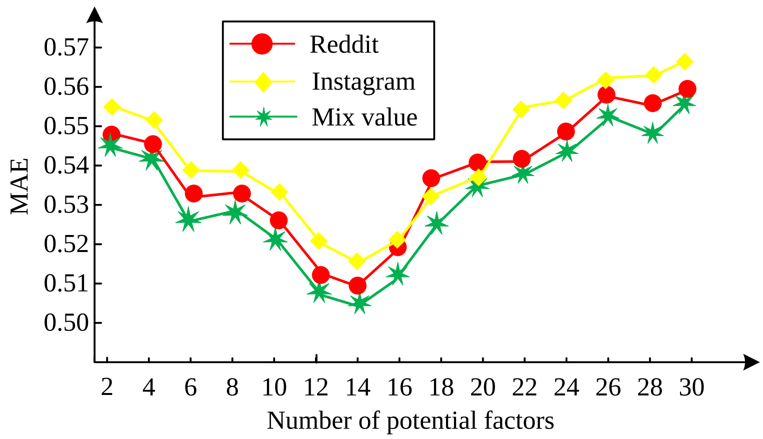
(b) Instagram dataset test result



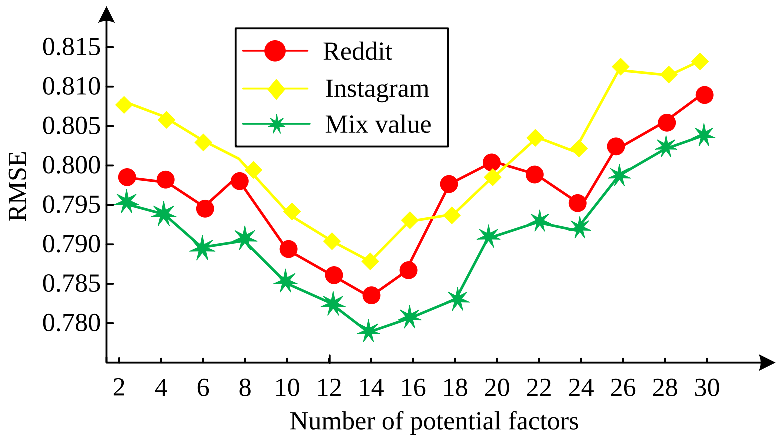
(c) Mixed dataset test result

Figure 8: The influence of embedding dimension s on RMSE

In Fig. 8 (a), the fluctuation range of the RMSE image is relatively small for the value of s. At s=96, the minimum RMSE value obtained is 0.75, and at s=256, it is increased by about 7% compared to the minimum value. In Fig. 8 (b), the maximum RMSE value obtained at s=256 is 0.819, and the minimum RMSE value obtained at s=96 is 0.756. The mixed dataset test results in Fig. 8 (c) show that the s value at the minimum RMSE is also 96. In the model parameter settings, s=96 is the optimal embedding dimension value. An excessive number of potential factors can render the model unduly complex, which may result in a reduction in model performance. To ascertain the optimal range of potential factor numbers, experiments are conducted with a range of factor numbers from 0 to 30, based on the parameter settings of other studies. The results of the change in the number of potential factors t with other parameters fixed are shown in Fig. 9.



(a) MAE value



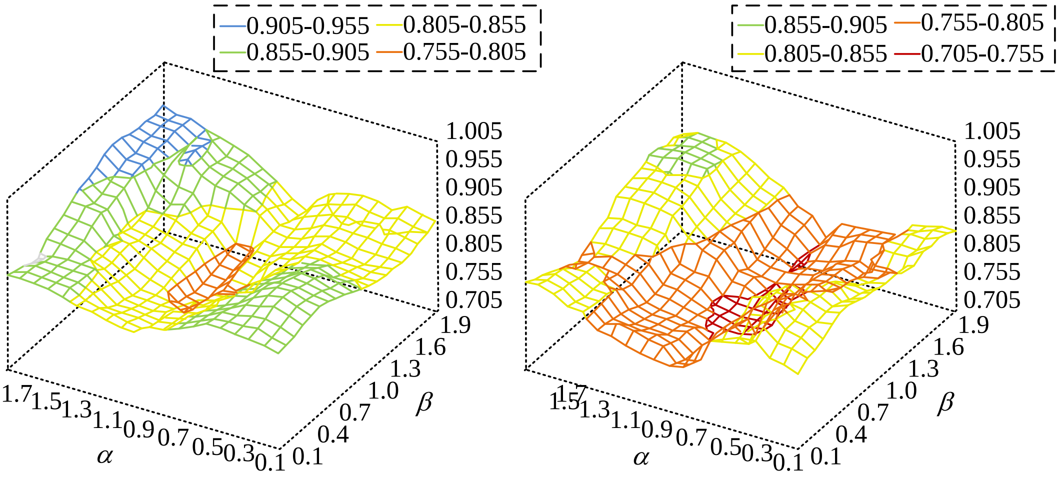
(b) RMSE value

Figure 9: The influence of the number of potential factors t on the experimental results

In Fig. 9 (a), when there are potential factors $t = 14$, the MAE values tested on Instagram, Reddit, and mixed datasets are the smallest, at 0.516, 0.509, and 0.506. In Fig. 9 (b), when the number of potential factors is $t = 14$, the RMSE values tested on Instagram, Reddit, and mixed datasets are the smallest, with values of 0.787, 0.783, and 0.778. Therefore, in the model parameter settings, the optimal value for the number of potential factors is 14.

3.3 Performance analysis of SIP recognition model

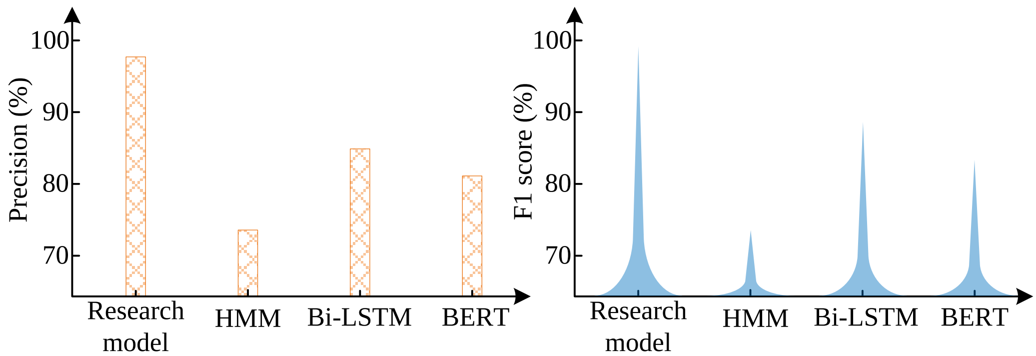
This study fixes the embedding dimension and the number of latent factors to 96 and 14, and tests the Instagram and Reddit datasets by simultaneously adjusting the fusion learning parameters, as shown in Fig. 10.



(a) RMSE for Reddit dataset testing (b) RMSE for Instagram dataset testing

Figure 10: The influence of fusion parameters α and β on experimental results

In Fig. 10 (a), the lowest RMSE value is in the middle of the orange area. At this time, the fusion learning parameters α and β have values around 1, and the interval of RMSE is $[0.755,0.805]$. When the values of α and β are close to 2, the blue part has the highest RMSE value. In 10 (b), when the values of α and β are around 1, the RMSE value in the red part in the middle is the lowest, and the range of RMSE is $[0.705,0.755]$. Therefore, when the fusion learning parameter value is 1, the test results on Instagram and Reddit datasets perform better. To further test the performance of the model, under the optimal parameters mentioned above, the study compares the model with other sensitive information review models on a mixed dataset, as shown in Fig. 11.



(a) The accuracy of different model recognition (b) F1 score recognized by different models

Figure 11: Comparison test results of SIP recognition models

In Fig. 11 (a), the testing accuracy of the research model is 97.3%, while the accuracy of HMM, Bi-LSTM, and BERT are 73.2%, 84.8%, and 80.7%, respectively. In Fig. 11 (b), the test F1 score of the research model is 97.8%, while the test F1 scores of HMM, Bi-LSTM, and BERT are 72.1%, 87.6%, and 82.3%, respectively. In summary, the SMSIR model based on BL-NE has better recognition accuracy and F1 score, demonstrating better performance.

4 Discussion

This study analyzed the embedding dimension, number of potential factors, and optimal parameters for fusion learning of the model, and compared the performance of the BL-NE-based SMSIR model. The results showed that when embedding dimension was s=96, the tests of MAE and RMSE on three datasets were equal to the minimum value. This result was similar to the research conclusion proposed by Zhang's team on determining the optimal embedding dimension in a network service recommendation model based on adaptive embedding representation [19]. The results indicated that the optimal embedding dimension obtained through testing could significantly improve the accuracy of model validation. In the test of the number of potential factors, when t was 14, the minimum values of MAE on the three datasets were 0.516, 0.509, and 0.506, and the minimum values of RMSE on the three datasets were 0.787, 0.783, and 0.778. This result was consistent with the conclusion of the Li et al. in designing a multi-source feature learning QoS prediction model based on joint deep networks [20]. In summary, the model had the highest accuracy in reviewing and identifying potential factors at the optimal parameter value, and performed better in HSNs with larger data volumes. Finally, the testing accuracy of the research model on mixed datasets was 97.3%, while the accuracy of HMM, Bi-LSTM, and BERT were 73.2%, 84.8%, and 80.7%, respectively. This was similar to the conclusion of Wang's team's dynamic heterogeneous information network embedding model based on Metapath proximity [21]. HMM can identify sensitive behaviors or content by analyzing changes in user behavior patterns, but the initial assumptions of HMM are relatively strict, which affects its accuracy. Bi-LSTM may be affected by training data, making it difficult to ensure recognition accuracy. BERT's bidirectional semantic understanding ability can capture subtle differences in text, but its insufficient generalization ability affects its recognition accuracy. The research method adopts network alignment of SIP, so identifying sensitive information with a locked publisher will have higher accuracy. This result indicates that the SMSIR model based on BL-NE is feasible and effective.

5 Conclusion

This study aimed to review and identify sensitive information posted by extreme individuals from a large amount of diverse information on social media platforms. By extracting and filtering effective information through MPRW in multi-platform HSNs, this study integrated network embedding and BL to map sensitive information and all nodes represented by users to a low-dimensional space. Afterwards, a fusion function was designed to predict the correlation score of multi-platform SIP, and the publisher and sensitive information were determined through the score. Experiments have shown that the SMSIR model based on BL-NE performed better than other models in reviewing and identifying sensitive information on multiple platforms, especially on platforms with larger data volumes. In practical applications, extreme individuals may not publish sensitive information on the large social media platforms mentioned in the study. In a social media platform with less interaction, the amount of data on this platform is relatively small and there may be fewer connections with other platforms. The difficulty of reviewing and identifying sensitive information in this situation will greatly increase. Therefore, future research will continue to investigate the sensitive information censorship of such small social media platforms.

References

[1] N. Steinfeld, "The disinformation warfare: How users use every means possible in the political battlefield on social media," Online Information Review, vol. 46, no. 7, pp. 1313-1334, 2022. https://doi.org/10.1108/OIR-05-2020-0197

[2] K. Omar, R. H. Sakr, and M. F. Alrahmawy, "An ensemble of CNNs with self-attention mechanism for DeepFake video detection," Neural Computing & Applications, vol. 36, no. 6, pp. 2749-2765, 2024. https://doi.org/10.1007/s00521-023-09196-3

[3] P. Katariya, V. Gupta, R. Arora, A. Kumar, S. Dhingra, Q. Xin, and J. Hemanth, "A deep neural network-based approach for fake news detection in regional language," International Journal of Web Information Systems, vol. 18, no. 5, pp. 286-309, 2022. https://doi.org/10.1108/IJWIS-02-2022-0036

[4] H. Chen, B. Jiang, and S. X. Ding, "A broad learning aided data-driven framework of fast fault diagnosis for high-speed trains," IEEE Intelligent Transportation Systems Magazine, vol. 13, no. 3, pp. 83-88, 2021. https://doi.org/10.1109/MITS.2019.2907629

[5] A. Mewada, and R. K. Dewang, "NRWalk2Vec-HIN: spammer group detection based on heterogeneous information network embedding over social media," Journal of Supercomputing, vol. 80, no. 2, pp. 1818-1851, 2024. https://doi.org/10.1007/s11227-023-05537-0

[6] S. Jeong, J. Park, and S. Lim, "mr2vec: Multiple role-based social network embedding," Pattern Recognition Letters, vol. 176, no. 12, pp. 140-146, 2023. https://doi.org/10.1016/j.patrec.2023.11.002

[7] M. M. H. Emon, G. M. M. D. N. Rabbani, and A. Nath, "Challenges and opportunities in the implementation of big dataanalytics in management information systems in Bangladesh," Acta Informatica Malaysia, vol. 7, no. 2, pp. 122-130, 2023. https://doi.org/10.26480/aim.02.2023.122.130

[8] C. Zhan, W. Jiang, F. Lin, S. Zhang, and B. Li, "A decomposition-ensemble broad learning system for aqi forecasting," Neural Computing & Applications, vol. 34, no. 21, pp. 18461-18472, 2022. https://doi.org/10.1007/s00521-022-07448-2

[9] G. Xiao, Y. Wei, H. Yao, W. Deng, J. Xu, and D. Pan, "Hierarchical broad learning system for hyperspectral image classification," IET Image Processing, vol. 16, no. 2, pp. 554-566, 2022. https://doi.org/10.1049/ipr2.12371

[10] C. Zeng, H. Liu, L. Lu, J. Chen, and Z. Zhou, "Shatter: searching heterogeneous combat network attack sequences through network embedding and reinforcement learning," IEEE Systems Journal, vol. 17, no. 3, pp. 4497-4508, 2023. https://doi.org/10.1109/JSYST.2022.3231346

[11] N. Ogino, "Resilient virtual network embedding ensuring connectivity under substrate node failures," IEICE Transactions on Communications, vol. 105, no. 5, pp. 557-568, 2022. https://doi.org/10.1587/transcom.2021EBP3122

[12] J. Shen, W. Dong, J. Wang, P. Zang, X. Shi, and Y.

Wang, "Time-sensitive satellite internet based on uniform content labels," Journal of Web Engineering, vol. 21, no. 2, pp. 523-544, 2022. https://doi.org/10.13052/jwe1540-9589.21217

[13] T. Zhang and H. Ma, "ClothNet: sensitive semantic segmentation network for fabric defect detection," Textile Research Journal, vol. 93, no. 1, pp. 103-115, 2023. https://doi.org/10.1177/00405175221114927

[14] D. Liang, F. Li, and Z. Xu, "A group-based FMEA approach with dynamic heterogeneous social network consensus reaching model for uncertain reliability assessment," The Journal of the Operational Research Society, vol. 74, no. 1, pp. 33-47, 2023. https://doi.org/10.1080/01605682.2021.2020694

[15] M. Wang, X. M. Li, X. Zhang, and Y. Zhang, "Hierarchical graph attention network with pseudo-metapath for skeleton-based action recognition," Neurocomputing, vol. 501, no. 8, pp. 822-833, 2022. https://doi.org/10.1016/j.neucom.2022.06.024

[16] Y. Wang, W. Wang, Z. Zhen, Q. Peng, P. Jiao, W. Liang, M. Shao, and Y. Sun, "Geometry interaction network alignment," Neurocomputing, vol. 501, no. 28, pp. 618-628, 2022. https://doi.org/10.1016/j.neucom.2022.06.077

[17] Y. Wang, Q. Peng, W. Wang, X. Guo, M. Shao, H. Liu, W. Liang, and L. Pan, "Network alignment enhanced via modeling heterogeneity of anchor nodes," Knowledge-Based Systems, vol. 250, no. 17, pp. 11-11, 2022. https://doi.org/10.1016/j.knosys.2022.109116

[18] K. C. Yuen, B. Tran, A. Anton, H. Hamidi, A. J. Costello, N. M. Corcoran, N. Lawrentschuk, N. Rainey, M. G. Semira, P. Gibbs, S. Mariathasan, S. Sandhu, and E. E. Kadel, "Molecular classification of hormone-sensitive and castration-resistant prostate cancer, using nonnegative matrix factorization molecular subtyping of primary and metastatic specimens," The Prostate, vol. 82, no. 9, pp. 993-1002, 2022. https://doi.org/10.1002/pros.24346

[19] H. Zhang, M. Wu, Q. Feng, and H. Li, "Aerqp: adaptive embedding representation-based qos prediction for web service recommendation," Journal of Supercomputing, vol. 80, no. 3, pp. 3042-3065, 2024. https://doi.org/10.1007/s11227-023-05582-9

[20] Y. Xia, D. Ding, Z. Chang, and F. Li, "Joint deep networks based multi-source feature learning for qos prediction," IEEE Transactions on Services Computing, vol. 15, no. 4, pp. 2314-2327, 2022. https://doi.org/10.1109/TSC.2021.3050178

[21] X. Wang, Y. Lu, C. Shi, R. Wang, P. Cui, and S. Mou, "Dynamic heterogeneous information network embedding with meta-path based proximity," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 3, pp. 1117-1132, 2020. https://doi.org/10.1109/TKDE.2020.2993870