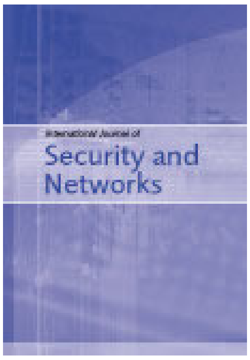

International Journal of Security and Networks

ISSN online: 1747-8413 - ISSN print: 1747-8405
https://www.inderscience.com/ijsn

A social network security user recommendation algorithm based on community user emotions

Huajin Liu, Chunhua Ju, Houyuan Zhang

DOI: 10.1504/IJSN.2023.10060574

Article History:

Received:	14 September 2023
Last revised:	15 September 2023
Accepted:	24 September 2023
Published online:	12 March 2024

A social network security user recommendation algorithm based on community user emotions

Huajin Liu*

School of Business Administration,
Zhejiang Gongshang University,
Hangzhou 310018, China
and
School of Network Communication,
Zhejiang YueXiu University,
Shaoxing 312000, China
Email: 20141026@zyufl.edu.cn
*Corresponding author

Chunhua Ju

School of Business Administration,
Zhejiang Gongshang University,
Hangzhou 310018, China
Email: jch@zjgsu.edu.cn

Houyuan Zhang

College of Humanities and Arts,
Ningbo University of Technology,
Ningbo 315211, China
Email: zhanghouyuan@nbut.edu.cn

Abstract: Social networks play a vital role in people's lives and work, but have problems with sparse data and cold start. This study establishes a social network model and innovatively improves the classic user interest point recommendation algorithm based on community information and user emotion. A sequential learning ranking algorithm is designed to simulate user preferences from a sequence of recommended objects and convert user ratings into ranking scores, combined with a network security dictionary, Node2vec method, and hot coding to capture network security vocabulary. This study also uses the heuristic firefly optimisation algorithm to solve the problem and confirms that community CU-SNR has good experimental results. The improved LDA algorithm is used to adjust the social media emotion data, and three real social network data sets verify the algorithm's performance. Numerical experiment results show that the algorithm simulation has a certain effect when facing social networks.

Keywords: community information; user characteristics; social networks; firefly algorithm; recommendation algorithm; linear discriminant analysis; LDA.

Reference to this paper should be made as follows: Liu, H., Ju, C. and Zhang, H. (2024) 'A social network security user recommendation algorithm based on community user emotions', *Int. J. Security and Networks*, Vol. 19, No. 1, pp.10–19.

Biographical notes: Huajin Liu is an Associate Professor. He is a PhD candidate at the Zhejiang Gongshang University. Now, he works in the Zhejiang Yuexiu University. His research interest includes computer simulation, big data mining, business intelligence related research, etc.

Chunhua Ju is a Professor and Doctoral Supervisor. Currently, he works in the Zhejiang Gongshang University who focuses on intelligent in-formation processing, data mining and e-commerce. And he won the award for 'New Century Excellent Talents in University' of China. In the past several years, he led more than six national projects.

Houyuan Zhang is a Professor and Master tutor. Currently, he works in the Ningbo Institute of Technology who focuses on online public opinion governance, international social media research, and cross-cultural communication. Thirteen academic papers have won international and domestic awards. In the past few years, he has presided over one national project and three provincial projects.

1 Introduction

In the progress of internet technology with a high speed, compared to the real society, social networks are very significant in people's work and life due to their advantages in time and space. Social networks have many users and massive data, making it very important for users to find other users who may become friends efficiently. However, users generally lack effective filters of surrounding information, and there are problems, such as inaccurate information delivery and low information utilisation in friend recommendations. How to design efficient friend recommendation algorithms has become an important research topic (Banati et al., 2014; Chiniah and Ghannoo, 2023). Therefore, as one way is quite effective in solving information overloaded, personalised user information recommendation methods have emerged, and their recognition in the academic community is also rapidly increasing. The user-personalised information recommendation technique based on community mining will be further applied to network information resource management, enterprise competitive intelligence, information dissemination, and other fields (Zhang et al., 2018).

Personalised recommendation technology has been widely used to recommend products, services, or activities to consumers, but consumers often do not exist alone. They are in various groups (Li et al., 2019). With the increasing attention of researchers to group activities, group recommendation systems have emerged for group activities such as movies, picnics, tourism, etc. (Abolghasemi et al., 2022). Online social networks have become a primary place where people live online, and group recommendation systems for online social networks are becoming a hot research topic whose field is recommendation. Due to different motivations for users to participate in the community, whether it is a self-motivated social choice or a social influence based on conformity, there will be significant differences in the community's organisational structure and decision-making mode (Valdez et al., 2018). In the process of social choice, people are more eager to form relationships with people with similar points. Homogeneity plays a role. People choose friends according to similar feature selection, which is the mechanism of people's self-active choice. Under the influence of social influence mechanisms, the social connections existing in the network will affect the characteristics of individual nodes (Ni et al., 2011).

Traditional recommendation algorithms are various, mainly comprising collaborative filtering and content-based recommendation algorithms. Moreover, the hybrid recommendation algorithm is also included (Gong et al., 2018). For example, linear discriminant analysis (LDA) is one algorithm of them. Furthermore, the collaborative filtering algorithm is widely used as a recommendation algorithm but still faces serious data sparsity and cold start problems (Gorripati and Vatsavayi, 2017). The content-based recommendation algorithm searches for similar items that have interacted with the user and then makes recommendations. Therefore, this approach requires effective feature extraction. The hybrid recommendation algorithm indicates mixed multiple recommendation methods to compensate for each other's shortcomings and achieve better recommendation results (Hu et al., 2018). Combining collaborative filtering technology with other technologies to overcome the cold start problem is the most common. In many application fields, the interaction information between many types of users and items will be recorded over time. Traditional recommendation algorithms that use entire historical data for recommendations cannot effectively capture users' short-term interests and preferences (Yan et al., 2021).

The classic social network user recommendation algorithms mainly include two types: based on user attribute feature similarity and based on the topological structure between users. Still, problems include inaccurate recommendation results or narrow recommendation ranges. The link prediction algorithm is widely used in friend recommendation algorithms (Zheng et al., 2022). The basic idea is to treat users in social networks as nodes in a network, use friend attributes to calculate the similarity between nodes and form new links to make friend recommendations. However, only a single network structure is considered in link prediction, resulting in low recommendation accuracy (Wang et al., 2019). At present, friend recommendation algorithms generally integrate diverse user attributes, such as user basic information, social relationships, geographical location, etc. to solve problems such as low recommendation accuracy and information overloaded. Providing personalised information recommendation services for users has become the development direction of many network platforms. However, we also clearly recognise that although personalised information recommendation has certain advantages in social network user recommendation, applying different personalised recommendation algorithms to information services will more or less face key issues, for instance, data sparsity, cold start, and system scalability can be included (Das et al., 2021).

This research aims to establish a social network model, innovatively improving the classic user interest point recommendation algorithm based on community information and user emotion (CU-SNR), using the sequential learning sorting algorithm for sorting and optimisation, and combining the network security lexicon, Node2vec method, and unique hot coding when obtaining and processing the network security vocabulary vector of social network users. This research proposes an innovative solution technology based on the improved firefly algorithm, using the improved LDA algorithm for adjusting social media emotional data, and verifies the algorithm performance through three real social network datasets.

In addition to the introduction section, the following sections are also included. The relevant models and the algorithms are detailed in the literature in Section 2. Section 3 introduces the model construction and algorithm optimisation in detail, and the algorithm is solved. Section 4 details the experimental operation steps and analyses the experimental results. Finally, in Section 5, we summarise the work of this study, propose the shortcomings, and make a prospect for future work.

2 Literature review

Inputting data is the first step when building a recommendation system. The cost of obtaining explicit scoring data increases during the process, and data sparsity becomes more severe. To some extent, the growth of internet date and users leads to lower recommendation accuracy and user satisfaction. Implicit user-project interaction data can be an ideal solution to this problem. There are three main collection methods for users to browse behaviour data: server-side, client-side, and server-client integration (Liu et al., 2019). The implicit interest score belonging to project users can be obtained by analysing and quantifying the collected browsing behaviour data. Based on users' various personalised needs and application scenarios, collaborative filtering and other recommendation algorithms are used to obtain personalised recommendation lists. Last but not least, the results will be sent to target users on the page through ranking lists, images, links, and other forms (Chen et al., 2020a). Based on multi-source heterogeneous data such as user and their browsing project attribute information, browsing behavior data, and user social network information, as well as their analysis and processing techniques.

Community information can substantially improve information filtering, filter from huge information sets at high speed, and offer retrieval users the information set that shows their interests and preferences with the highest correlation (Jalali and Hosseini, 2021). Scholars have made a comparison of the results of user browsing time preference analysis with the help of explicit user ratings and discovered that the time users spend reading their favourite newspapers is more than the one spent on regular articles, indicating that user browsing time can reflect user interest preferences effectively (Li and Chen, 2016). Scholars have applied collaborative filtering methods which are based on browsing time to Usenet news, further verifying that printing, saving, and adding bookmarks, etc. which are browsing behaviours that can show user interests and preferences to make up for the shortcomings of explicit feedback rating data (Zhao et al., 2016). For the recommendation system, the preference feedback, which belongs to individual users' browsing actions linked to the display feedback mark to get user preferences, is the main practical application that belongs to browsing behaviour in the early. It is a simple model in the early period based on the user's collaborative filtering recommendation (Salakhutdinov and Mnih, 2007). The internet develops rapidly, users have become increasingly more, and the prominent problem is information overload. The accuracy and stability of recommendations based solely on explicit feedback have declined, and the significance and need for implicit feedback data have increased (Weimer et al., 2007). For instance, browsing behaviour in personalised recommendation models is included. Implicit user project interaction data can solve this problem (Liu and Yang, 2008).

In user-based collaborative recommendation, the system requires calculating the similarity between users to recommend similar users. However, the large amount of historical data generated by frequent website visits will lead to insufficient online computing performance in online recommendation algorithms. Some scholars have developed one collaborative filtering algorithm based on items that can be used to solve the problem that the performance of online computing and the quality of recommendations cannot be considered (Shi et al., 2013a). Because of the relatively static relationships between items, this algorithm calculates item similarities instead of user similarities, which avoids performance issues in online computing, thus achieving similar effects for recommending (Shi et al., 2013b). Although the latent factor model cannot be used alone in session recommendations because of the insufficiency of user information, it still has modelling capabilities that neighbourhood models do not possess. The neighbourhood model calculates the most similar items rated by users, a local optimisation problem that does not consider all items globally. The latent factor model can represent items as a whole. Still, it does not perform well in detecting the strongly correlated material between several items having a stronger correlation, which shows the ascendancy of the neighbourhood model precisely (Wu et al., 2017). Therefore, some scholars have proposed using the advantages of neighbourhood and latent factor models to improve prediction accuracy. For calculating item similarity in neighbourhood models, this model proposes a more accurate neighbourhood model (Yang et al., 2021). At the same time, the model incorporates implicit data into the model and expands the model. Moreover, it also merges the latent factor model and neighbourhood model and utilises explicit and implicit user feedback to obtain more effective recommendation results (Hsu et al., 2018).

Meanwhile, other researchers have proposed similar algorithms to solve link prediction problems in social networks. For instance, singular value decomposition (SVD), probabilistic matrix factorisation (PMF), heats, non-negative matrix factorisation (NMF), Jaccard, common neighbours (CN), and preferential attachment (PA) are included (Zhang, 2020; Wang et al., 2017; Zhou et al., 2010; Nasiri et al., 2023; Kumar et al., 2022).

3 Algorithm optimisation

3.1 *Classic recommendation algorithm*

The collaborative filtering recommendation algorithm has some main steps. It needs to construct a user data model, calculate the similarity measure of user preferences, select and predict the nearest neighbour users, and select the item with the highest prediction score as the recommended item to give feedback to the target users (Huang et al., 2019).

Calculating the preference similarity between users is critical to collaborative filtering. Common methods contain Cosine similarity, modified Cosine similarity, and Pearson similarity calculation method. Firstly, the modified Cosine similarity is introduced. That is, the difference between different user scoring standards will bring about a large error in Cosine similarity. The modified Cosine similarity can reduce the error by decreasing the average mark of users on the item. Equation (1) is available for reference. *Sim* (*a*, *b*) represents the preference similarity of users *a* and *b*. $R_{a,i}$ shows the evaluation of project *i* from the user *a*, and $\overline{R_a}$ shows the average rating of all projects from the user *a*, $R_{b,i}$ shows the evaluation of project *i* from the user *b*, and $\overline{R_b}$ shows the average rating of all projects from the user *b*.

$$sim(a,b)=\frac{\sum i\in I\left(R_{a,i}-\overline{R_a}\right)*\left(R_{b,i}-\overline{R_b}\right)}{\sqrt{\sum i\in I\left(R_{a,i}-\overline{R_a}\right)^2}*\sqrt{\sum i\in I\left(R_{b,i}-\overline{R_b}\right)^2}} \quad (1)$$

When calculating Pearson similarity, the similarity between different users can be obtained based on their common project preferences as shown in equation (2). $v_{a,j}$ shows the evaluation of project *j* by the user *a*, and $\overline{v_a}$ shows the average rating of all projects from the user *a*, $v_{b,j}$ shows the evaluation of project *j* from the user *b*, and $\overline{v_b}$ shows the average rating of all projects from the user *b*, while the correlation between user U_a and U_b is between –1 and 1 (Davtalab and Alesheikh, 2021). In addition, users with a similarity greater than the threshold are also selected as similar users to the target user. Then, the target user's rating of unrated items can be predicted based on similar neighbour users' rating that belongs to the target item, as shown in equation (3).

$$sim(a,b)=\frac{\sum\left(v_{a,j}-\overline{v_a}\right)\left(v_{b,j}-\overline{v_b}\right)}{\sqrt{\sum_j\left(v_{a,j}-\overline{v_a}\right)^2}\sqrt{\sum_j\left(v_{b,j}-\overline{v_b}\right)^2}} \quad (2)$$

$$P_{ai}=\overline{R_a}+\frac{\sum b\in N_a sim(a,b)*\left(R_{b,i}-\overline{R_b}\right)}{\sum b\in N_a sim(a,b)} \quad (3)$$

The classic recommendation algorithm based on user features first constructs one rating matrix about user items. Meanwhile, the rating data which belongs to the user set $U=\{u_1, u_2, \cdots, u_m\}$ on the project set $I=\{I_1, I_2, \cdots, I_n\}$ is converted into the user project rating matrix $R(m, n)$, and *rui* represents users' rating for project *i*. The figure of the project rating can be divided into several levels, such as 0~5 or 0~1, indicating some user's behaviour, such as adding people to a shopping cart, browsing, etc. (Wang and Cao, 2020). To recommend something for the target users and obtain neighbour user ratings, the formula $P(\cdot)$, which is used to predict the rating for the project *i* evaluated by the user *u*, can be shown in equation (4).

$$P(\cdot)=\frac{v\in\sum_{Nu}sim(u,v)*\left(Rvi-\overline{Rv}\right)+\overline{Ru}*\sum_{Nu}sim(u,v)}{\sum_{Nu}sim(u,v)} \quad (4)$$

3.2 Model construction and algorithm optimisation

In social networks, the connection between users is generally represented by constructing a relationship graph, where a node represents every user $u_i \in V$, and their interactions are indicated by edges $(u_i, u_j) \in E$. The community structure in social networks implies that user nodes can be divided into subsets $C=\{C_1, C_2, \cdots, C_k\}$, making nodes C_j in the same subset tightly connected, making the connections between subgroups relatively sparse. Existing research mainly focuses on disjoint community structures and ensures that the belonging of each node is only one community. Users' forwarding, collecting, and commenting on information can be considered positive response behaviour in social networks. Therefore, the social network user set can be defined as $U=\{a_1, a_2, a_3, \cdots, a_n\}$ and the set of social network information is defined as $I=\{i_1, i_2, i_3, \cdots, i_k\}$.

Let $I(\cdot)$ express the interest level $\frac{\overline{Luj}}{\overline{Lu}}$ which belongs to the user u for the users from the project attribute set A and $\overline{Luj}$ is the average figure of all scores for the sub-attribute *j* evaluated by the user *u* and $\overline{Luj}$ shows the average figure of all scoring items for the user *u*, so several sub-attributes of design items and the similarity of user preferences for project sub-attributes are shown in equation (5).

$$sim_P(u,v)=\frac{\sum_{j=1}^{n}\left(Puj-\overline{Pu}\right)\left(Pvj-\overline{Pv}\right)}{\sqrt{\sum_{j=1}^{n}\left(Puj-\overline{pu}\right)^2}\sqrt{\sum_{j=1}^{n}\left(Pvj-\overline{Pv}\right)^2}} \quad (5)$$

Then, this study designed a sequential learning sorting algorithm. It is essential to convert the user's rating values for different recommended objects into sorting scores to simulate the information users prefer from the sequence of recommended objects sorted by the user's rating value. The conversion from scoring values to ranking values is achieved using the first-place probability (Wu et al., 2021). From a probability perspective, the first-place probability shows the probability that a recommendation object can be ranked first in the recommendation sequence according to all recommendation objects. The first-place probability and its variants can usually convert the scoring value into one certain range of probability values. It is assumed that the score for the recommended audience *i* evaluated by the user *u* is R_{ui} so that the first-place probability and objective function $H(\cdot)$ of the user *u* recommending the audience *i* are shown in equations (6), and (7), respectively.

$$P\left(R_{ui}\right)=\frac{e^{R_{ui}}}{\sum_{k=1}^{N}e^{R_{uk}}} \quad (6)$$

$$H(\cdot)=\sum_{i=1}^{N}\frac{e^{R_{ui}}}{\sum_{k=1}^{N}e^{R_{uk}}}*\left(\sum_{u=1}^{M}\|U_u\|_F^2+\sum_{i=1}^{N}\|V_i\|_F^2\right) \tag{7}$$

Considering the great similarity between the user's emotional change and memory forgetting, the forgetting curve describes the emotional change of users. Because of the time forgetting function $f(t)$ of the user, an improved method for calculating user interest preferences for sub-attributes can be further obtained, where the similarity between user and item sub-attributes for users u and v is shown in equation (8).

$$sim(u,v)=\frac{\sum_{j=1}^{n}(P(T)-\bar{u})*(P(T)-\bar{v})}{\sqrt{\sum_{j=1}^{n}(P(T)-\bar{u})^2}\sqrt{\sum_{j=1}^{n}(P(T)-\bar{v})^2}} \tag{8}$$

Sentiment analysis mainly constructs an emotion dictionary by labelling the polarity and intensity of words in the text, then classifies the text to calculate its emotion value. However, the emotional dictionaries of comment texts in different fields may differ. Therefore, the vocabulary constructed in this article includes two parts: a basic sentiment dictionary and an extended sentiment dictionary composed of sentiment words supplemented based on comment texts. Based on this dictionary, the emotional values of positive and negative sentiment in the text can be effectively calculated. Using the interpoint mutual information (PMI) algorithm [as shown in equation (9)], based on the seed words with positive emotion and the ones with negative emotion as the benchmark words, subtract the interpoint mutual information of candidate emotion words and judge the user's emotional tendency by the difference size. Within it $P(\omega_1)$ indicates the frequency with which ω_1 appears in the corpus, $P(\omega_1 \cap \omega_2)$ indicates the probability of ω_1 and ω_2 in one sentence, a larger $PMI(\omega_1, \omega_2)$-value indicates a more pronounced positive emotion.

$$PMI(\omega_1,\omega_2)=\log_2\left[\frac{P(\omega_1\cap\omega_2)}{P(\omega_1)*P(\omega_2)}\right] \tag{9}$$

User comments on social networks reflect their attention to text attributes. Extracting user interests based on the comment text and conducting word frequency statistics can obtain user attention points (Chen et al., 2020b). Therefore, the word frequency ratio of e ach attribute is used as the weight of the recommendation index, and sentiment analysis is performed on the comment text of each user attribute to calculate the user's sentiment value towards the user's information. The final social network user recommendation algorithm based on CU-SNR in this article is shown in equation (10).

$$R(\cdot)=\sum_{k=1}^{t}\omega_k*\frac{\sum_{j=1}^{n}(P(T)-\bar{u})*(P(T)-\bar{v})}{\sqrt{\sum_{j=1}^{n}(P(T)-\bar{u})^2}\sqrt{\sum_{j=1}^{n}(P(T)-\bar{v})^2}}*\frac{T_k-F_k}{T_k+F_k} \tag{10}$$

Within it, the attribute weight of the k^{th} user is represented by ω_k, and T_k is the positive emotion of the k^{th} user and F_k is the negative emotion of the k^{th} user, and t represents the quantity of texts. The social media data in this model includes relational data, text data, and basic information data. Organise the text and basic information data of all users, extract and select a network security vocabulary, and form a network security vocabulary vector $\boldsymbol{V}$. According to the vector $\boldsymbol{V}$, it can perform unique hot coding on nodes to form multidimensional vectors. If the corresponding network security words appear in the social text and basic information of the user node, they are marked as 1 in the corresponding dimension and 0 if not. This vector is the network security feature vector of the user node. Splice the feature vectors of each user node with the feature vectors generated using Node2vec to get the user nodes' final representation vectors. Use this node to represent vectors for downstream tasks to complete network security users' classification, recognition, and recommendation.

3.3 Solving algorithm

This study uses a heuristic firefly optimisation algorithm for solving, inspired by the behaviour of fireflies in nature. The authors (Hashem and Hassanein, 2019) consider that fireflies are generally considered neutral to simplify the characteristics of algorithm application, so the attraction of fireflies is Unisex neutral. In this motion, the dim firefly will move towards the bright firefly, and each firefly's brightness can represent the solutions' quality. Commonly, the quality of the solution is directly proportional to the value of the objective function. When calculating the initial solution of the firefly algorithm, if considering that there is N member population in the D-dimensional environment, then every solution can be expressed as $E_i = \{E_w, E_x, E_v, E_z\}$. The attraction of fireflies is calculated based on Cartesian distance, shown in equation (11).

$$L'(i)=L(i)+\beta_0*e^{-\gamma*r^2}*(x_i-x_j)+\alpha*\varepsilon_i \tag{11}$$

Within it, $L(i)$ shows the new position of the firefly i, and α shows the step scale factor, and ε_i shows the random factor generated by a uniform distribution from 0 to 1. First and foremost, if the initial population of fireflies exists, it is a random population. And compare two types of fireflies, (i.e., two solutions), with lower brightness fireflies (weaker solutions) moving towards higher brightness fireflies (better solutions). Then, update the positions of all fireflies and continue these steps until completing the comparison of all fireflies, as shown in Algorithm 1.

Algorithm 1 Solution technology based on improved firefly algorithm

1: Generate initial population of fireflies $i = [1, N] \in \{w, x, y, z\}$ randomly

2: Calculate the objective function $f(w, x, y, z)$ of the brightness of firefly nodes

3: Define attractiveness parameters $\{w, x, y, z\}$

```
Iterate on the period
For i ∈ [1, N], j ∈ [1, i]
    Define the movement of the firefly i towards the
    firefly j in the social network environment
    Calculate the distance d between the user node i and j
    Update calculation of user node attractiveness e^{−γ*r^2}
End For
Sort firefly nodes by influence
Return {w, x, y, z}
```

4 Simulation examples

4.1 Experimental design and data description

This study evaluates the proposed algorithm here. According to the approach of references (Li et al., 2014; Breese et al., 2013), three real social network datasets were used in the experiment (as shown in Table 1). This study used Python software to select 30 influential users as the initial nodes of social networks, based on recent hot public opinion event keywords 'COVID-19' and 30 hot comments user nodes, the above variables as the individual and organisation of media image crisis sample has strong representative, respectively climbed the sina weibo, zhihu and Facebook Chinese and English user data set as the basis of the experimental simulation data (climb time for November 24, 2021–March 15, 2022). This study treats each user as a node, using edges between nodes to represent user relationships. This study selected 30 influential users and their friends list as the initial nodes of social networks to generate simple social networks. The research team implemented the proposed algorithm and the relevant algorithms for comparison on Tensorflow1.5.1. The experiment was conducted in groups, and the data were divided into ten groups by cross-validation. That is, the dataset was divided into ten equal parts. One set of data was selected to play the role of the test set at a time, and the other groups as the training set, and the final average value was taken. And save all data in CSV format in the MySQL database for data processing. For each social network dataset, 10% of each user rating data was selected randomly as the test set using the Rapidminer data mining tool, and the residual 90% of user data was used to become the training set. The experiments were implemented in the Python framework, in addition to the conventional Numpy, Scipy, Pandas, Matplotlib, and Theano packages using the Surprise package. And on two Linux operating system servers (Intel Xeon processor (34 GHz) 64 GB memory), each with a 6-core CPU, two NVIDIA Titan X GPU, and 100 GB RAM. Since the results of the experiment's public opinion control model may differ in each run, the evaluation results are set as the average value after 500 iterations, and the standard deviation of the operation is 1.415.

Emotional analysis in social media estimates the total number of positive emotional information related to information over time. This study constructed sentiment word banks, polarity word banks, and negative words, respectively, and adjusted the social media sentiment data of the listed companies using the improved LDA algorithm shown in Algorithm 2. The sentiment lexicon is based on various sources, including the sentiment vocabulary list published by China National Knowledge Infrastructure, the 'detailed dictionary of commonly used commendatory and derogatory words', the 'student commendatory and derogatory words dictionary', the 'commendatory words dictionary', and the 'derogatory words dictionary'. It also removes the low frequency of sentiment words. It adds online and spoken sentiment words, including 4,637 commendatory words and 5,139 derogatory words, and divides the constructed emotional lexicon into five levels based on the frequency of vocabulary usage, ranging from the simplest version to the complete version, denoted as KW1-KW5.

Algorithm 2 Improved LDA algorithm

```
Input: CNRDS Corpus document
Output: Document Theme Emotional Score
for topic z and emotion m
    Extract multiple distributions Ø_{z,m}
    for document d, and emotion j
        Extract multiple distributions φ_d, θ_dj
        for every sentence s and word w_{d,n} of sentence s
            Extract binomial distribution m_s, π_{n−1}
            if x_n = 0
                Extract multiple variable topics Z_n and
                word w_n
            else if x_n = 1
                Extract topic Z_{n−1} and word w_{n−1} from
                LDA distribution with parameter δ
            end if
        end for
    end for
end for
Return to document topic sentiment score
```

Furthermore, this study also identified a polar lexicon in the emotional lexicon (which includes some highly polar emotional words), especially some derogatory words. When identifying viewpoint sentences, as long as the above words appear, the polarity of the viewpoint sentence is determined as the polarity of the word (in negative sentences, the opposite is taken). This article refers to this type of lexicon as a polar lexicon to distinguish between large emotional lexicons, which include 158 positive and 281 negative words.

Table 1 Social network dataset

Network serial number	*Social network name*	*Type*	*Number of nodes*	*Number of node boundaries*	*Average degree*	*Node average path*	*Cluster coefficient*
1	Facebook network	Directed	217,430	2,371,849	29.385	5.41	0.526
		Directed	241,727	1,495,043	19.405	4.38	0.327
2	Sina Weibo network	Directed	294,821	3,294,393	61.293	4.15	0.602
		Directed	184,935	2,859,383	32.192	4.43	0.538
3	Zhihu network	Directed	58,373	273,843	49.288	5.34	0.532
		Directed	34,825	248,155	48.594	4.39	0.495

Table 2 AUC mean and standard deviation of different recommendation methods in different social networks

Dataset name	*Indicator name*	*SVD*	*PMF*	*Heats*	*NMF*	*Jaccard*	*CN*	*PA*	*CU-SNR*	*P_Value*
Facebook network	Mean	0.58	0.78	0.40	0.68	0.57	0.57	0.54	*0.83*	0.0000
	SD	0.04	0.02	0.01	0.02	0.02	0.02	0.01	*0.02*	-
Sina Weibo network	Mean	0.76	0.40	0.89	0.89	0.53	0.56	0.58	*0.92*	0.0000
	SD	0.02	0.01	0.01	0.02	0.02	0.03	0.03	*0.01*	-
Zhihu network	Mean	0.73	0.88	0.83	0.40	0.63	0.68	0.67	*0.92*	0.0000
	SD	0.02	0.01	0.01	0.01	0.01	0.02	0.02	*0.01*	-

Note: The values displayed in bold indicate that the corresponding algorithm performs well.

Table 3 The average accuracy and standard deviation of different methods in different social networks

Dataset name	*Indicator name*	*SVD*	*PMF*	*Heats*	*NMF*	*Jaccard*	*CN*	*PA*	*CU-SNR*	*P_Value*
Facebook network	Mean	0.12	0.37	0.50	0.00	0.02	0.00	0.03	*0.70*	0.0000
	SD	0.01	0.02	0.02	0.00	0.01	0.00	0.01	*0.00*	-
Sina Weibo network	Mean	0.80	0.73	0.65	0.79	0.14	0.08	0.18	*0.71*	0.0000
	SD	0.09	0.05	0.06	0.57	0.71	0.00	0.00	*0.00*	-
Zhihu network	Mean	0.00	0.30	0.13	0.16	0.03	0.05	0.01	*0.54*	0.0001
	SD	0.00	0.11	0.05	0.05	0.01	0.02	0.01	*0.01*	-

Note: The values displayed in bold indicate that the corresponding algorithm performs well.

The recommendation question in social networks is usually considered an assignment to realise the binary classification. When evaluating a binary classification task with two categories, the confusion matrix includes not only true positive (TP) and true negative (TN) but also false positive (FP) and false negative (FN). Additionally, in reference (Karunasingha and Santhusitha, 2022; Yu and Li, 2010), two precision functions are used: mean absolute error (MAE) and root-mean-square deviation (RMSE). The specific calculation methods are shown in equations (12), and (13), respectively. Within it, f_i represents the predicted value, y_i represents the true value.

$$MAE - \frac{1}{N}\sum_{i=1}^{N}\left|f_i - y_i\right| \tag{12}$$

$$RMSE - \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(observed_i - predicted_i\right)^2} \tag{13}$$

4.2 Experimental results

The CU-SNR algorithm proposed in this paper was used in numerical experiments, and other benchmark algorithms, such as SVD, PMF, heats, NMF, Jaccard, CN, and PA, were employed. Table 2 reports the area under the curve and its standard error results of the proposed CU-SNR algorithm when facing the dataset of the real social network. This study found the proposed CU-SNR algorithm had better experimental results in datasets. Table 3 reports the average accuracy and standard deviation results obtained by the CU-SNR algorithm proposed in this paper and other benchmark algorithms in real social network datasets. The results illustrate a high average accuracy in all experimental datasets of social networks, which belongs to the CU-SNR algorithm proposed in this research.

Table 4 reports the MAE and RMSE values belonging to the proposed CU-SNR algorithm and other benchmark algorithms in different social network topologies. As the MAE and RMSE values increase, the accuracy of the prediction optimisation algorithm decreases. This indicates that the proposed algorithm in this paper has a higher recommendation efficiency. According to Table 5, the proposed CU-SNR algorithm is generally superior to other benchmark recommendation methods. This is because the proposed CU-SNR algorithm uses a sequential learning sorting algorithm to optimise the recommendation process, ranking fast recommended objects based on the score of user information, and can respond quickly and recommend in real-time. It can also minimise the loss of the recommendation system.

Table 4 Comparison results of algorithms in power grid topology structure

	Indicator	*SVD*	*PMF*	*Heats*	*NMF*
Actual value	MAE	0.6383	0.6738	0.6239	0.5371
	RMSE	0.7362	0.6988	0.6582	0.5481
Optimal value	MAE	0.5353	0.6074	0.6248	0.5743
	RMSE	0.6353	0.6428	0.6739	0.5938
	Indicator	*Jaccard*	*CN*	*PA*	*CU-SNR*
Actual value	MAE	0.5382	0.5644	0.4272	*0.2371*
	RMSE	0.5463	0.5738	0.4371	*0.2492*
Optimal value	MAE	0.4739	0.5192	0.4472	*0.1281*
	RMSE	0.5291	0.5332	0.4738	*0.2455*

Note: The italic part indicates that this algorithm method is relatively optimal under this parameter condition.

Table 5 Comparison of algorithm single run time

Dataset name	*SVD*	*PMF*	*Heats*	*NMF*
Facebook network	0.1419	0.5121	5.1410	3.2419
Sina Weibo network	0.6378	4.2218	6.4833	9.0412
Zhihu network	0.8124	1.0422	7.1923	5.5407
Dataset number	*Jaccard*	*CN*	*PA*	*CU-SNR*
Facebook network	0.1538	0.2759	0.4634	*0.1324*
Sina Weibo network	0.4269	0.5846	0.6504	*0.1726*
Zhihu network	0.2435	0.4763	0.3638	*0.1449*

Note: The italic part indicates that this algorithm method is relatively optimal under this parameter condition.

5 Summary

The main contributions of this paper are summarised as follows.

First, by building a social network model, this paper innovatively improves the classic user interest point recommendation algorithm based on CU-SNR. It uses a sequential learning ranking algorithm for optimisation. It combines a network security dictionary, Node2vec method, and unique hot coding when obtaining and processing network security term vectors.

Second, this paper proposes a solution technology based on the improved Firefly algorithm, uses the improved LDA algorithm to adjust the emotional data of social media, and verifies the algorithm's performance through three real social network data.

Third, for calculating item similarity in the neighbourhood model, a more accurate domain model is proposed in the model presented in this paper. At the same time, the model incorporates implicit data into the model and extends the model. In addition, the potential factor and domain models are integrated and explicit and implicit user feedback is used to obtain more effective recommendation results, providing a reference for subsequent scholars' research.

This research uses the social network model to improve the classic safe user recommendation algorithm based on community information and user sentiment, uses the sequential learning sorting algorithm to sort and optimise, and then uses the Firefly algorithm to solve the proposed CU-SNR algorithm. This study also used three datasets obtained from real crawlers to solve the proposed CU-SNR algorithm in a social network data environment. The results achieved by numerical experiments show that the algorithm model proposed in this research can recommend user information efficiently and accurately when facing social networks. Despite some significant findings mentioned above, this study still has certain limitations, some of which may point the way for further research in the future. Firstly, the uncertainty of nodes and edges can be considered simultaneously in social network models to enhance further the proposed algorithm's ability to resist uncertainty. Secondly, it is possible to explore integrating time information, location information organiser information, etc. into the framework that belongs to social network recommendation algorithms to obtain better performance which belongs to recommendation algorithms and better solve cold start issues. In this way, the efficiency and accuracy of the algorithm model proposed in this research can improve. Finally, cutting-edge technologies such as deep Autoencoder can make social network information recommendations more accurately and efficiently.

Acknowledgements

This work was supported by the National Social Science Fund of China, 'A Study on the Public Opinion Guidance Mechanism of Network Ideological Security in the 5G Background' (Grant No. 21BXW071).

References

Abolghasemi, R., Engelstad, P., Herrera-Viedma, E. and Yazidi, A. (2022) 'A personality-aware group recommendation system based on pairwise preferences', *Information Sciences*, Vol. 595, No. 2022, pp.1–17.

Banati, H., Mehta, S. and Bajaj, M. (2014) 'Social behaviour based metrics to enhance collaborative filtering', *Journal of Computer Information Systems*, Vol. 6, No. 2014, pp.217–226.

Breese, J.S., Heckerman, D. and Kadie, C. (2013) 'Empirical analysis of predictive algorithms for collaborative filtering', *Uncertainty in Artificial Intelligence*, Vol. 98, No. 7, pp.43–52.

Chen, J.R., Zhao, C.X., Uliji, and Chen, L.F. (2020a) 'Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering', *Complex & Intelligent Systems*, Vol. 6, No. 1, pp.147–156.

Chen, R., Chang, Y.S., Hua, Q.Y., Gao, Q.L., Ji, X. and Wang, B. (2020b) 'An enhanced social matrix factorization model for recommendation based on social networks using social interaction factors', *Multimedia Tools and Applications*, Vol. 79, Nos. 19–20, pp.14147–14177.

Chiniah, A. and Ghannoo, F. (2023) 'A multi-theory model to evaluate new factors influencing information security compliance', *International Journal of Security and Networks*, Vol. 18, No. 1, pp.19–29.

Das, J., Majumder, S., Gupta, P. and Datta, S. (2021) 'Scalable recommendations using decomposition techniques based on Voronoi diagrams', *Information Processing & Management*, Vol. 58, No. 2021.

Davtalab, M. and Alesheikh, A.A. (2021) 'A POI recommendation approach integrating social spatio-temporal information into probabilistic matrix factorization', *Knowledge and Information Systems*, Vol. 63, No. 1, pp.65–85.

Gong, J.B., Chen, S., Gao, X.X., Song, Y.Q. and Wang, S.L. (2018) 'Integrating LDA into the weighted average method for semantic friend recommendation', *6th CCF Academic Conference on Big Data, CCF Big Data 2018*, pp.427–441.

Gorripati, S.K. and Vatsavayi, V.K. (2017) 'Community-based collaborative filtering to alleviate the cold-start and sparsity problems', *International Journal of Applied Engineering Research*, Vol. 12, No. 15, pp.5022–5030.

Hashem, M. and Hassanein, A.S. (2019) 'Jaw fracture classification using meta heuristic firefly algorithm with multi-layered associative neural networks', *Cluster Computing-the Journal of Networks Software Tools and Applications*, Vol. 22, No. 3, pp.S7079–S7086.

Hsu, C.C., Yeh, M.Y. and Lin, S.D. (2018) 'A general framework for implicit and explicit social recommendation', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 30, No. 12, pp.2228–2241.

Hu, J.J., Liu, L.Z., Zhang, C.Y., He, J.L. and Hu, C.Z., (2018) 'Hybrid recommendation algorithm based on latent factor model and personal rank', *Journal of Internet Technology*, Vol. 19, No. 3, pp.919–926.

Huang, L., Tan, W.S. and Sun, Y. (2019) 'Collaborative recommendation algorithm based on probabilistic matrix factorization in probabilistic latent semantic analysis', *Multimedia Tools and Applications*, Vol. 78, No. 7, pp.8711–8722.

Jalali, S. and Hosseini, M. (2021) 'Social collaborative filtering using local dynamic overlapping community detection', *Journal of Supercomputing*, Vol. 77, No. 10, pp.11786–11806.

Karunasingha, D. and Santhusitha, K. (2022) 'Root mean square error or mean absolute error? Use their ratio as well', *Information Sciences*, Vol. 585, No. 2022, pp.609–629.

Kumar, M., Mishra, S. and Biswas, B. (2022) 'Features fusion based link prediction in dynamic neworks (R)', *Journal of Computational Science*, Vol. 57, No. 2022.

Li, G. and Chen, Q. (2016) 'Exploiting explicit and implicit feedback for personalized ranking', *Mathematical Problems in Engineering*, Vol. 2016, No. 2016, pp.1–11.

Li, H., Li, H.N., Zhang, S., Zhong, Z.M. and Cheng, J. (2019)' Intelligent learning system based on personalized recommendation technology', *Neural Computing & Applications*, Vol. 31, No. 9, pp.4455–4462.

Li, R., Lin, H. and Yan, J. (2014) 'Mining latent semantic on user-tag-item for personalized music recommendation', *Journal of Computer Research &* Development, Vol. 51, No. 10, pp.2270–2276.

Liu, N.N. and Yang, Q. (2008) 'EigenRank: a ranking-oriented approach to collaborative filtering', *31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM SIGIR 2008*, pp.83–90.

Liu, X., Zhou, Y.J. and Wang, Z.R. (2019) 'Acquisition method of users' browsing behavior preference based on the fusion of social network link and theme model', *Journal of Intelligent & Fuzzy Systems*, Vol. 37, No. 1, pp.493–508.

Nasiri, E., Berahmand, K. and Li, Y.F. (2023) 'Robust graph regularization nonnegative matrix factorization for link prediction in attributed networks', *Multimedia Tools and Applications*, Vol. 82, No. 3, pp.3745–3768.

Ni, S.J., Weng, W.G. and Zhang, H. (2011) 'Modeling the effects of social impact on epidemic spreading in complex networks', *Physica A-Statistical Mechanics and Its Applications*, Vol. 390, Nos. 23–24, pp.4528–4534.

Salakhutdinov, R. and Mnih, A. (2007) 'Probabilistic matrix factorization', *Advances in Neural Information Processing Systems*, Vol. 17, No. 5, pp.1257–1264.

Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M. and Hanjalic, A. (2013a) 'XCLiMF: optimizing expected reciprocal rank for data with multiple levels of relevance', *7th ACM Conference on Recommender Systems, RecSys 2013*.

Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M., Oliver, N., Hanjalic, A. (2013b) 'CLiMF: Collaborative less-is-more filtering', *23rd International Joint Conference on Artificial Intelligence, IJCAI 2013*, pp.3077–3081.

Valdez, A.C., Brell, J., Schaar, A.K. and Ziefle, M. (2018) 'The diversity of why: a meta-analytical study of usage motivation in enterprise social networks', *Universal Access in the Information Society*, Vol. 17, No. 3, pp.549–566.

Wang, R.Q., Cheng, H.K., Jiang, Y.L. and Lou, J.G. (2019) 'A novel matrix factorization model for recommendation with LOD-based semantic similarity measure', *Expert Systems with Applications*, Vol. 123, No. 2019, pp.70–81.

Wang, S.J. and Cao, L.B. (2020) 'Inferring implicit rules by learning explicit and hidden item dependency', *IEEE Transactions on Systems Man Cybernetics-Systems*, Vol. 50, No. 3, pp.935–946.

Wang, Y.S., Liu, F.B., Xia, S.T. and Wu, J. (2017) 'Link sign prediction by variational Bayesian probabilistic matrix factorization with student-t prior', *Information Sciences*, Vol. 405, No. 2017, pp.175–189.

Weimer, M., Karatzoglou, A., Viet L.Q. and Smola, A. (2007) 'CoFiRANK maximum margin matrix factorization for collaborative ranking', *Proceedings of the 21th International Conference on Neural Information Processing Systems*, Curran Associates Inc, Vancouver, pp.79–86.

Wu, H., Yue, K., Hsu, C.H., Zhao, Y.J., Zhang, B.B. and Zhang, G.Y. (2017) 'Deviation-based neighborhood model for context-aware QoS prediction of cloud and loT services', *Future Generation Computer Systems-The International Journal of Escience*, Vol. 76, No. 2017, pp.550–560.

Wu, L., Sun, P.J., Hong, R.C., Ge, Y. and Wang, M. (2021) 'Collaborative neural social recommendation', *IEEE Transactions on Systems Man Cybernetics-Systems*, Vol. 51, No. 1, pp.464–476.

Yan, J., Zeng, Q.T. and Zhang, F.Q. (2021) 'Summary of recommendation algorithm research', *Journal of Physics: Conference Series*, Vol. 1754, No. 1, p.012224.

Yang, X., Xie, G. and Li, X. (2021) 'Identifying influential spreaders in complex networks using neighbourhood coreness and path diversity', *International Journal of Security and Networks*, Vol. 16, No. 3, pp.174–182.

Yu, H. and Li, Z.Y. (2010) 'A collaborative filtering method based on the forgetting curve', 2010 *International Conference on Web Information Systems and Mining, WISM 2010*, pp.183–187.

Zhang, C.J. (2020) 'An orthogonal matching pursuit algorithm based on singular value decomposition', *Circuits Systems and Signal Processing*, Vol. 39, No. 1, pp.492–501.

Zhang, H., Ge, D. and Zhang, S. (2018) 'Hybrid recommendation system based on semantic interest community and trusted neighbors', *Multimedia Tools and Applications*, Vol. 77, No. 4, pp.4187–4202.

Zhao, L.L., Lu, Z.Q., Pan, S.J. and Yang, Q. (2016) 'Matrix factorization+ for movie recommendation', *25th International Joint Conference on Artificial Intelligence, IJCAI 2016*, pp.307–315.

Zheng, J.X., Qin, Z.F., Wang, S.G. and Li, D.Y. (2022) 'Attention-based explainable friend link prediction with heterogeneous context information', *Information Sciences*, Vol. 597, No. 2022, pp.211–229.

Zhou, T., Kuscsik, Z., Liu, J.G., Medo, M., Wakeling, J.R. and Zhang, Y.C. (2010) 'Solving the apparent diversity-accuracy dilemma of recommender systems', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 10, pp.4511–4515.