Community Discovery Algorithm for Social Networks Based on Parallel Recommendation

Zhai Jian-Li1,*, Lin Wei-Xuan2

1Huali College Guangdong University of Technology, Guangdong Guangzhou, 511325, China
2Guangzhou Huali Science and Technology Vocational College, Guangzhou, 511325, China

*Corresponding author E-mail: zhaijianli1@126.com

Abstract: The distribution of user community in social network has the problem of similar feature distribution, which leads to the high viscosity between community and community, and it is difficult to mine community features. A social network community discovery algorithm is proposed based on parallel recommendation. The fuzzy decision model of social network community discovery is established by taking the user behavior value, user consumption value and loyalty of social network as independent variables. The relationship between the interaction degree and the recommendation effect in the social network is analyzed by extracting the characteristic quantity of the community association attribute in the social network, and the community group recommendation in the social network is carried out by using the parallel recommendation algorithm, and the number of visits between users is analyzed. The number of messages is used as the weighted weight coefficient, the difference factor is introduced to evaluate the influence of the community, the synchronous label of the social network community is established, and the community feature mining and parallel recommendation are realized according to the label location. The simulation results show that, the proposed method has good accuracy in community discovery, high accuracy in community attribute feature mining, and low cost of the algorithm. It has obvious advantages compared with similar algorithms.

1. Introduction
With the development of network information technology and the emergence of a large number of social networks, people establish social groups and communities in social networks, make the relationship between people networking. With the rapid development of information technology and the popularity of the Internet. Social networks have launched new life and vitality, and a large number of social network-based services and applications have grown rapidly[1]. In China, Tencent QQ, WeChat and Douban.com represent the rapid development of social networks. Social networking is developed and designed by using a variety of social network software. For example, the social network service platform is constructed by BlogWIKI Web2.0 and other core Web2.0 applications. The key to the development and design of social network lies in the optimization design of community discovery algorithm, and the design of community algorithm through social network is the key to the
development and design of social network. The recommendation ability and adaptive recognition ability of the network are improved[2].

The community of social network refers to the community of some specific groups connected together according to a certain relationship[3]. According to the characteristics of multi-level tasks, the community model of social network is established, and the community is divided according to the topological structure of the network. Build a conceptual model in human-computer interaction to rebuild communities and build virtual relationships through social networks, in which users can publish text, pictures, videos and other information via the network. The rise and popularity of online communities has changed people's traditional habits and ways of making friends, and it has changed the way people access information and purchase behavior to some extent[4]. Because the online community of social networks uses a one-to-many mode of dissemination, when users post content, all friends can see it. This has resulted in a form of information dissemination, that is accessing to information through friends[5]. At present, the online community discovery algorithms for social network mainly include collaborative filtering recommendation algorithm and similarity feature mining algorithm. The corresponding community feature mining algorithm is used to predict the users' scores for the unevaluated items, and then a group model is established to recommend and discover the community, so as to improve the community's ability of group recommendation and collaborative filtering. However, the traditional algorithm has the problems of too much computing overhead and poor real-time detection in the community[6].

In this paper, a social network community discovery algorithm based on parallel recommendation is proposed. The fuzzy decision model of social network community discovery is established based on the social network user behavior value, user consumption value and loyalty as independent variables. The relationship between the interaction degree and the recommendation effect in the social network is analyzed by extracting the characteristic quantity of the community association attribute in the social network, and the community group recommendation in the social network is carried out by using the parallel recommendation algorithm, and the number of visits between users is analyzed. The number of messages is used as the weighted weight coefficient, the difference factor is introduced to evaluate the influence of the community, and the synchronous label of the social network community is established. Community feature mining and parallel recommendation are realized according to tag location. Finally, simulation experiments are carried out to demonstrate the application performance of this method in improving the social network online community discovery and optimization identification.

2. A fuzzy decision model for social network community discovery

2.1. Analysis of constraint variables found in social networking communities

In order to realize community discovery and recognition of social networks, a parallel recommendation algorithm is used to make fuzzy recommendation decisions for users and communities, and a statistical analysis model is used according to the community functions and attributes of social networks. The fuzzy prediction model of social network community discovery is established[7]. Firstly, the community distribution model of social network is given as shown in figure 1.
The architecture of the social network in Figure 1 is described by an undirected graph \( G = (V, E) \), where node \( V \) is any node in the network, namely \( v \in V \), and the decentralized Decentralized MSNs data transfer protocol is used for information exchange and storage in the network structure\(^{8}\). The community fuzzy information recognition of social network in \( m \)-dimensional distributed feature space is carried out, and the community function score data set of social network is obtained:

\[
X(n) = \{x(n), x(n+\tau), \ldots, x(n+(m-1)\tau)\} \quad n = 1, 2, \ldots, N
\]

(1)

Wherein, \( \tau \) represents the users of the large potential semantic model, \( m \) is the dimension of user vector. Based on the modeling, the feature vectors of all users are extracted, and the phase trajectory evolution analysis is carried out in the high-dimensional feature space, so as to value the user behavior of the community network. As independent variables, the potential variable model of social network can be expressed as:

\[
X = [s_1, s_2, \ldots, s_k]
\]

(2)

Wherein, \( K = N - (m-1)\tau \) represents the vector related to user interest, \( \tau \) is the delay, \( m \) is the layer number of semantic ontology attribute of information, \( s_i = (x_{1i}, x_{2i}, \ldots, x_{(m-1)i})^T \) is called the feature vector set of each data tuple. In the fuzzy information distributed structure reorganization model, the group module is adopted. The difference degree is used to adjust the group scoring model, and the high-order statistical analysis model of community attribute distribution is constructed, and the fuzzy feature extraction of the community is carried out\(^{9}\).

2.2. Fuzzy decision function found in social network communities

Use \( I = \{i_1, i_2, \ldots, i_n\} \) and \( U = \{u_1, u_2, \ldots, u_m\} \) show item set and user set, by using collaborative filtering algorithm, the phase space reconstruction of big data fuzzy information in social network community distribution is carried out. The models are given as follows:

\[
H(S) = -\sum_{i=1}^{n} P_i(s_i) \log_2 P_i(s_i)
\]

(3)

\[
H(Q) = -\sum_{j=1}^{m} P_j(q_j) \log_2 P_j(q_j)
\]

(4)

Wherein, \( P_i(s_i) \) denotes the probability that the social network community distribution attribute concept set \( s_i \) appears in the clustering area \( S \). Similarly, \( P_j(q_j) \) denotes the probability of the social network community distribution ontology feature concept set \( q_j \) appears in the clustering.
region \( Q \). Under the condition of consistent similarity, the average mutual information in community attribute feature \( S \) satisfying fuzzy information clustering condition is calculated as follows:

\[
I(Q,S) = H(Q) - H(Q|S)
\]  

(5)

Where

\[
H(Q|S_i) = - \sum_j \left[ \frac{p_{ij}(s_j,q_j)}{p_i(s_j)} \right] \log_2 \left[ \frac{p_{ij}(s_j,q_j)}{p_i(s_j)} \right]
\]

(6)

It represents the classification attribute set satisfying the condition \( P \). The method of average mutual information fusion is used, according to the average score given, the high dimension characteristic quantity of the community is extracted, and the expression of fuzzy decision function found by the community is obtained:

\[
\begin{align*}
\frac{dx}{dt} &= -\sigma x + \sigma y \\
\frac{dy}{dt} &= -xz + rx - y \\
\frac{dz}{dt} &= xy - bz
\end{align*}
\]

(7)

Each variable in the formula is dimensionless after reduction, each quantity in the formula is dimensionless after reduction, and \( t \) is the sampling time of fuzzy data, \( x, y, z \) mean the community attribute partition variable, \( \sigma, r, b \) are fuzzy constraint parameters.

3. Optimization of social network community discovery algorithm

3.1. Feature extraction of community association attributes in social networks

Based on the establishment of fuzzy decision making model of social network community discovery, we design and optimize community discovery algorithm in social network. This paper proposes a community recommendation algorithm based on parallel recommendation. Calculate the community group \( G_j \) score set for item \( i \) as follows:

\[
R_{nm} = \|X_{q_{nm}} - X_n\|^{(m)} = \min_{j=N_{ij} \cdots N_{iN}} \|X_n - X_j\| = \sqrt{\sum_{j=0}^{n} (X_{q_{nm}} - X_{nm})^2}
\]

(8)

By adjusting the initial group model, the project recommendation optimization values of each user in the community in the social network are expressed as follows:

\[
R_{(m+1)k} = \|X_{q_{k}} - X_{nm}^{(m+1)}\| = \sqrt{\sum_{j=0}^{n} (X_{q_{k}} - X_{nm}^{(m+1)})^2} = \sqrt{\|X_{q_{k}} - X_{nm}^{(m)}\|^2 + \sum_{j=0}^{n} (X_{q_{nm}} - X_{nm})^2}
\]

(9)

If the \( R_{(m+1)k} \) is much larger than \( R_{(m+1)k} \), it means that the attribute disturbance of the social network community does not affect each other, and can be used as the fuzzy clustering center. According to this characteristic, the community discovery criteria for social networks are given as follows:

Criteria (1):

\[
\sqrt{\frac{R_{(m+1)k}^2 - R_{nm}^2}{R_{(m+2)k}^2}} = \frac{X_{q_{(m+1)r}} - X_{r+mod}}{R_{(m+1)r}} \geq R_{ol}
\]

(10)

Criteria (2):

\[
\frac{R_{(m+1)k}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i - \frac{1}{N} \sum_{i=1}^{N} x_i}} > A_{ol}
\]

(11)

Where, \( R_{ol} \) is the evaluation threshold for overall satisfaction, \( A_{ol} \) is fuzzy decision coefficient, set \( A_{ol} = 2 \). Extract the feature quantity of the community association attribute in the social network:
Extract the community distribution fuzzy recommendation decision quantity, as follows:

\[ R_i = \{X_1, X_2, \cdots, X_n\} \]

The set of association rule vectors recommended by social networks is:

\[ R^T_i R_i = \{X_1, X_2, \cdots, X_n\} \{X_1, X_2, \cdots, X_n\}^T \]

Based on the above analysis, for each user, the relationship between the interaction degree and the recommendation effect is analyzed, and the parallel recommendation algorithm is used to recommend the community group in the social network.

### 3.2. Parallel recommendation implementation of Social network Community Discovery

Parallel recommendation algorithm is used to recommend community group in social network, singular decomposition method is used to decompose the characteristic value of community information in social network, it is:

\[ R^T_i R_i = V \sum_i V^T_i \]

The output eigenvalues of fuzzy information extracted by users distributed in each community are obtained as follows:

\[ R^T_i R_i = V \sum_i V^T_i \]
\[ R_i = \{X_{d1}, X_{d2}, \cdots, X_{dm}\}^T \]

\[ R^T_i R_i = \{X_{d1}, X_{d2}, \cdots, X_{dm}\} \{X_{d1}, X_{d2}, \cdots, X_{dm}\}^T \]

By means of the number of visits and the number of messages between users as the weighted weights, the clustering categories that should belong to the users are calculated, and the test set \( V = [V_1, V_2, \cdots, V_n] \in R^{m \times n} \) of the community recommendation fuzzy information of social networks is obtained, it is orthogonal, means \( V V^T = I_n \), \( \sum = diag(\sigma^2, \sigma^2, \cdots, \sigma^2) \in R^{m \times m} \), introducing latent variables to construct association rule vector set of community feature distribution, and introducing difference factor to evaluate the influence of community, the fuzzy eigenvector set \( R^T R \) of community discovery satisfies inter-class equilibrium, and the order of fusion degree of community discovery is shown as follows:

\[ \sigma^2 > \sigma^2 > \sigma^2 > \cdots > \sigma^2 > \sigma^2 \]

It can be seen that the parallel recommendation algorithm is used to realize the community discovery and fuzzy evaluation decision of social network. The realization steps are described as follows:

1) Compute the similarity feature vectors between households of community network users \( \tilde{X}(i, n_i) \), in the social network, the fixed information is clustered to get the dispersion degree of the community partition \( \hat{\delta}_w \):

\[ \hat{\delta}_w = \sum_{i=1}^{n_i} \frac{1}{n_i} \sum_{i=1}^{n_i} \left[ \left( \tilde{X}_k - \bar{m}_i \right) \left( \tilde{X}_k - \bar{m}_i \right)^T \right] \]

2) According to the clustering attribute of social network nodes, the center vector is determined, and the dispersion matrix of fuzzy clustering of community distribution is obtained

\[ S_w = \sum_{i=1}^{n_i} p_i \left( \bar{m}_i - \bar{m} \right) \left( \bar{m}_i - \bar{m} \right)^T \]

3) By using the method of decision statistical regression analysis, the characteristic attribute set of community information is calculated, and the output characteristic quantity is obtained:

\[ j(\bar{X}_j) = \frac{y_j^T S_w y_j}{\lambda_j}, j = 1, 2, \cdots, I \]
4) Calculate the weighted matrix of community recommendations for Social Networks 
\[ W = [y_1, y_2, \cdots, y_d] \].

5) According to label location, community feature mining and parallel recommendation are realized. The results of community discovery feature mining are as follows: 
\[ \bar{X}' = W^\top \bar{X} \].

4. Simulation experiment and performance analysis

In order to test the application performance of this method in realizing the social network community discovery, the simulation experiment is carried out, the experiment is implemented with Matlab simulation tool, and the BookCrossing data set of the DeepWeb database in the social network soybean valve net is used in the experiment. For the test set, the test set contains 59456 users of Book-Crossing social network, the sample set of community attribute feature distribution is OAEI. The average value of 2000 users of the test set is taken, and the individual user model is established, and the number of users in each group is calculated. The correlation coefficient of the community is 0.23, the root mean square error between the quantitative estimated value of community discovery and the actual evaluation value is 0.014. Under the different weight coefficient, the community of social network is set. The correlation coefficient is shown in table 1.

Table 1. Community correlation coefficients under different weights

| w1   | Community correlation coefficients |
|------|-----------------------------------|
| 0.21 | 0.432                             |
| 0.25 | 0.554                             |
| 0.32 | 0.332                             |
| 0.45 | 0.425                             |
| ...  | ...                               |
| 0.56 | 0.532                             |
| 1    | 0.823                             |

According to the above simulation environment and parameter setting, the parallel recommendation of social network and the simulation analysis of community discovery are carried out, and the behavior characteristics of the users of community network are sampled. The time domain waveform of the sample is shown in figure 2.

Fig. 2 Time domain waveform of behavior feature sampling for community network users

Fig. 3 Community distribution clustering results of social networks

The community user behavior value, user consumption value and loyalty index collected from Figure 2 are taken as independent variables, and the mining of Community Association attribute attributes is carried out to achieve the clustering of community distribution characteristics, and the clustering results are shown in Figure 3.

Figure 3 shows that, this method is used to process the community distribution clustering of social networks, the clustering of data is better, and the community discovery ability is improved. On the
basis of this, different methods are used to test the accuracy of community discovery. The result of the test is shown in figure 4. The analysis figure 4 shows that the output accuracy of this method for community discovery of social networks is high, the stability of the parallel recommendation to the community is better, and the computational overhead is low.

![Fig. 4 Comparison of predictive accuracy of community findings](image)

5. Conclusions
By the design of online community discovery algorithm of social network, the latent knowledge contained in social network is extracted, and the online identification and optimal recommendation of social network community are realized. This paper presents a social network community based on parallel recommendation. Discovery algorithms. Based on the user behavior value of the community network, as independent variables, the user consumption value and loyalty are used as independent variables to establish a fuzzy decision model of community discovery in social network, extract the characteristic quantity of community association attribute in social network, and analyze the degree of interaction and the effect of recommendation in the group. The parallel recommendation algorithm is used to recommend the community groups in social networks, the number of messages is used as the weight coefficient, the difference factor is introduced to evaluate the influence of the community, the synchronous label of the social network community is established, and the community feature mining and parallel recommendation are realized according to the tag location. Simulation results show that the output accuracy of this method for social network community discovery is high, the stability of the parallel recommendation to the community is better, and the overall performance is superior.

References:
[1] LI Yuezhi, ZHU Yuanyuan, ZHONG Ming. k-core filtered influence maximization algorithms in social networks. Journal of Computer Applications, 2018, 38(2): 464-470.
[2] CAO J X, DONG D, XU S, et al. A k-core based algorithm for influence maximization in social networks[J]. Chinese Journal of Computers, 2015, 38(2):238-248.
[3] WANG C, CHEN W, WANG Y. Scalable influence maximization for independent cascade model in large-scale social networks[J]. Data Mining & Knowledge Discovery, 2012, 25(3):545-576
[4] LI G L, CHU Y P, FENG J H, et al. Influence maximization on multiple social networks[J]. Chinese Journal of Computers, 2016, 39(4):643-656.
[5] NING Xueli, LUO Yonglong, XING Kai, ZHENG Xiaoyao. Frequent location privacy-preserving algorithm based on geosocial network. Journal of Computer Applications, 2018, 38(3): 688-692.
[6] YE A Y, LI Y C,MA J F, et al. Location privacy-preserving method of k-anonymous based-on service similarity[J]. Journal on Communications, 2014, 35(11):162-169.
[7] WANG L, MENG X F, GUO S N. Preservation of implicit privacy in spatio-temporal data publication[J]. Journal of Software, 2016, 27(8):1922-1933.

[8] WU Zheng, YU Hongtao, LIU Shuxin, ZHU Yuhang. User identification across multiple social networks based on information entropy. Journal of Computer Applications, 2017, 37(8): 2374-2380.

[9] YE N, ZHAO L, DONG L, et al. User identification based on multiple attribute decision making in social networks[J]. China Communications, 2013, 10(12):37-49.