Computers, Materials & Continua CMC, vol.60, no.2, pp.659-674, 2019

Collaborative Filtering Recommendation Algorithm Based on Multi-Relationship Social Network

Sheng Bin[1, *], Gengxin Sun[1], Ning Cao[2], Jinming Qiu[2], Zhiyong Zheng[3], Guohua Yang[4], Hongyan Zhao[5], Meng Jiang[6] and Lina Xu[7]

Abstract: Recommendation system is one of the most common applications in the field of big data. The traditional collaborative filtering recommendation algorithm is directly based on user-item rating matrix. However, when there are huge amounts of user and commodities data, the efficiency of the algorithm will be significantly reduced. Aiming at the problem, a collaborative filtering recommendation algorithm based on multi-relational social networks is proposed. The algorithm divides the multi-relational social networks based on the multi-subnet complex network model into communities by using information dissemination method, which divides the users with similar degree into a community. Then the user-item rating matrix is constructed by choosing the k-nearest neighbor set of users within the community, in this case, the collaborative filtering algorithm is used for recommendation. Thus, the execution efficiency of the algorithm is improved without reducing the accuracy of recommendation.

Keywords: Big data, recommendation algorithm, social network, recommendation efficiency.

1 Introduction

As an effective means to solve the problem of information overload [Liu (2016)], recommendation system is also one of the most common applications in the field of large data. According to the user's history information, hobby and other characteristics, the recommendation system uses knowledge detection technology to filter users' interesting information and products, so as to realize personalized recommendation. At present, the most commonly used recommendation system technology is recommendation algorithm

[1] School of Data Science and Software Engineering, Qingdao University, Qingdao, 266071, China.

[2] College of Information Engineering, Sanming University, Sanming, 365004, China.

[3] Department of Information Engineering, Fujian College of Water Conservancy and Electric Power, Yongan, 366000, China.

[4] School of Internet of Things and Technology, Wuxi Vocational Institute of Commerce, Wuxi, 214153, China.

[5] College of Information Engineering, Shandong Yingcai University, Jinan, 250104, China.

[6] School of Electronic and Information, Shandong Modern University, Jinan, 250104, China.

[7] School of Computer Science, University College Dublin, Dublin, Ireland.

[*] Corresponding Author: Sheng Bin. Email: binsheng@qdu.edu.cn.

based on collaborative filtering [Dong and Ke (2018)]. The kind of algorithm calculates the similarity of users or commodities based on the user-item rating matrix, and then recommends the commodities that users may need. In addition, the real meaning of big data is not big but multiple sources [Zhang, Li and Tang (2017)]. How to find out the inner relationship of multi-source data in the process of fusion through scientific analysis, and extract a unified, richer and more valuable information than a single data source, which is the real significance of using large data. Therefore, integrating the social relationship among users into the recommendation system can calculate the similarity of users more accurately. To a certain extent, it also solves the problems of sparse data and cold start [Margaris, Vassilakis and Georgiadis (2018)] in traditional recommendation algorithm based on collaborative filtering. Actually, there are many relationships among users of social networks, each of which has a different impact on recommendation. Therefore, simply introducing a certain social relationship will inevitably affect the accuracy of recommendation results.

In view of the above-mentioned facts, a collaborative filtering algorithm based on multi relational social networks is proposed in the article. Firstly, the multi-relational social network is constructed by the multi-subnet complex network model [Sun and Bin (2018)]. Then, the algorithm divides the multi-relational social networks based on the multi-subnet complex network model into communities by using information dissemination method, which divides the users with similar degree into a community. Next, the k-nearest neighbor set is composed of the K users with the highest similarity with the recommended users in the community, which is used to construct the user-item rating matrix. Finally, the collaborative filtering algorithm is used for recommendation. Because of the sparse relationship between community organizations, the similarity between members of different community organizations is weak. Therefore, it is not necessary to analyze the user-item rating matrix constructed by the whole data set in the recommendation process, thus improving the execution efficiency of the algorithm without reducing the accuracy of recommendation.

2 Related works

Collaborative filtering technology is mainly divided into memory-based [Zhao (2013)] collaborative filtering and model-based [Jiang, Qian and Shen (2015)] collaborative filtering. Both of the two methods are based on user-item rating matrix. However, with the increase of the number of users and commodities, the user-item rating matrix will become larger and larger, resulting in the performance of collaborative filtering algorithm is decreasing. In order to improve the efficiency of recommendation algorithm, Roh et al. [Roh, Oh and Han (2003)] use hierarchical clustering algorithm to cluster users and commodities, then it is recommended based on the weighted summation of root nodes to leaf nodes. Zhang et al. [Zhang, Ding and Jiang (2005)] used fuzzy clustering method to classify users who like similar items and items that liked by the same user before recommending. Santhini et al. [Santhini, Balamurugan and Govindaraj (2015)] put forward the basic idea of collaborative filtering recommendation for big data by clustering method, which improves the efficiency of recommendation algorithm to a certain extent. However, clustering technology relies on the user-item rating matrix as the original data source to reflect personal preferences. It does not take into account the social relationships such as friendship and trust between users, nor does it consider the existence of implicit relationships between the scored commodities or items. Ignoring these important relationships will inevitably lead to the distortion of recommendation results. Accordingly, the introduction of social networks into the recommendation system to improve the rating matrix of the social recommendation system [Li, Li, Chen et al. (2018)] came into being. Chang et al. [Chang, Diaz and Hung (2015)] fused the trust relationship in social networks in the recommendation algorithm, and predicted the score by using trust between users as an alternative to the vacancy values in the user-item matrix. Pan et al. [Pan, Zhang and Wang (2016)] alleviated the problem of the sparse data and cold start by using the trust relationship in social network, which proved that social network can greatly improve the performance of recommendation system. As a typical complex network [Sun and Bin (2017)], social networks have the characteristics of community structure, which can divide the network members into groups with close relationship within the group and sparse relationship between groups [Jiang, Sun and Bin (2018)]. Wang et al. [Wang, Li and Zhang (2016)] used weighted label dissemination algorithm to divide the citation cooperative network into communities, and then used recommendation algorithm to recommend academic papers in the same community. Zhang et al. [Zhang, Zhen and Zhang (2018)] mapped the user-item rating matrix to the user similarity network and combined with the community in the network to recommend the sorted Top-n commodities to users, which improved the accuracy and efficiency of the recommendation. Although the existing recommendation algorithms based on community partitioning in complex networks have made some progress in the accuracy and efficiency of recommendation, they are all based on single-relational social networks, in which only one relationship exists. In fact, social network is actually a multi-relationship network, for example, there are user trust, similar interests and other relationships among users of social network. These explicit or implicit relationships determine the interaction characteristics between users together, which have different impact on the recommendation results. Therefore, the social network introduced into the recommendation system should be a multi-relational complex network with heterogeneous nodes, while the traditional complex network model cannot describe these complex relationships well, nor can it describe the interaction between heterogeneous nodes. Based on that, the article constructs a multi-relationship social network based on the multi-subnet complex network model which can describe the multi-relationship between different types of individuals, and divides the social structure of the network by information dissemination method. When predicting users' rating, the similarity of users' rating is calculated according to the degree of users' preference, and then the k-nearest neighbor set is selected for the recommended users in the community by combining the revised similarity of users' rating and the similarity of user social interaction, thus greatly reducing the computational complexity of the recommendation algorithm.

3 Construction of multi-relationship social network based on multi-subnet complex network model

In the recommendation system, its elements can be divided into two sets: user and commodity. Suppose there are m users and n commodities, the user set U can be expressed as $U = \{u_1, u_2, \ldots, u_m\}$, and the commodity set I can be expressed as

$I = \{i_1, i_2, \ldots, i_n\}$. Based on that, the user's rating matrix R for commodities is composed of $m * n$ elements $r_{11}, r_{12}, \ldots, r_{mn}$, which is represented by $[r_{ui}]_{m\times n}$, where r_{ui} represents the user $u's$ rating for the product i, and the user-item rating matrix is shown in Fig. 1.

	i_1	i_2	i_3	i_4	i_5
u_1	3		5		4
u_2		2		3	
u_3			2		1
u_4		4			1
u_5	2	5			4

Figure 1: User-item rating matrix

In reality, the recommendation process is usually accompanied by the integration of social information, such as the user's behavior of inquiring about friends' shopping opinions, the sharing of shopping information between friends and so on. Actually, if there is a trust relationship between users, the similarity of their interest preferences will be significantly higher than that of strangers [Guo, Li and Zhao (2016)], that is, if one makes a high rating on a commodity, then another user is likely to show a high rating on the commodity.

The trust relationship in social networks can be represented by the trust-rating matrix T, which is expressed as $[t_{uv}]_{m\times m}$, where t_{uv} represents the user $u's$ rating on user v, it reflects the degree of trust between users. The trust-rating matrix is shown in Fig. 2.

	u_1	u_2	u_3	u_4	u_5
u_1			1		
u_2					3
u_3		1			
u_4					2
u_5	2				

Figure 2: Trust-rating Matrix

In order to effectively integrate the social feature, the user-item rating network and the user trust relationship network are combined into multi-relational social network based on the multi-subnet complex network model.

3.1 Construction of user-item rating network and user trust relationship network

The multi-subnet complex network model is a quaternion satisfying the following conditions: $G = (V, E, R, F)$, where $v = \{v_1, v_2, \ldots, v_m\}$ denotes the set of nodes in the network, $m = |V|$ is the number of nodes in the set; $E = \{\langle v_h, v_l \rangle | v_h, v_l \in V, 1 \leq h, l \leq m\} \subseteq V \times V$ is the set of edges between nodes; and $R = R_1 \times \ldots \times R_i \times \ldots \times R_n = \{(r_1, \ldots, r_i, \ldots, r_n) \ | r_i \in R_i, 1 \leq i \leq n\}$, R_i denotes the set of interaction relations among nodes, r_i is one of them, n is the total number of interaction relations, and if $|R| \geq 2$, the network is a multi-relationship network; $F: E \overset{\varphi}{\rightarrow} R$ denotes that the mapping of the edge set E is projected by a function φ to find a unique corresponding mapping in F, which represents the type of relationship on edge.

Single-relationship complex network is a special case of multi-subnet complex network. By using multi-subnet complex network model, user-item rating network R_G and user trust relationship network T_G are constructed based on user-item rating matrix and trust-rating matrix respectively, and threshold R and T are set. If the user $u's$ rating of commodity i is greater than or equal to the threshold R, it is considered that the commodity satisfies the user's preference.. Similarly, if the user $u's$ trust rating of another user v is greater than or equal to the threshold T, it is considered that there is a trust relationship between users. In the user-item rating network R_G, each node symbolizes a user or a commodity, and $R = R_1 = \{r_1\}$ indicates the preference relationship between users and commodities. Similarly, in the user trust relationship network T_G, each node represents a user, and $R = R_2 = \{r_2\}$ represents the trust relationship between users.

The network is constructed according to the user-item rating matrix and trust-rating matrix given in Fig. 1 and Fig. 2, where the threshold R and T are both set to 2, then the network topology is constructed as shown in Fig. 3 and Fig. 4.

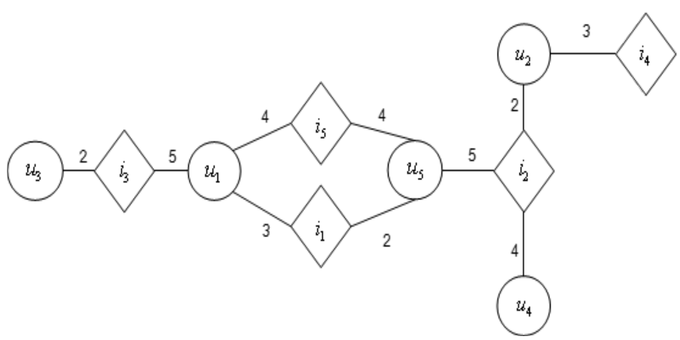

Figure 3: User-item rating network

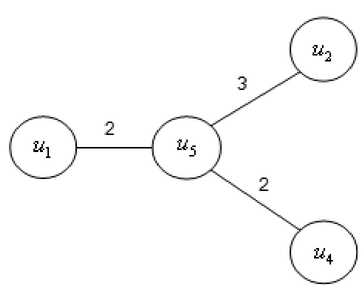

Figure 4: User trust relationship network

3.2 Construction of multi-relationship complex networks

Multi-subnet complex network can compose two or more complex networks by subnet loading operation. In the process of loading, the network is mapped to vector space, and the relationship between nodes is mapped to multidimensional vector in vector space. Therefore, the process of subnet loading can be transformed into the base transformation of vector space, and the change of the relationship between the related nodes is accomplished by the correlation operation of the space vector. Due to the inheritance of the nature of each subnet, the dimension of the relationship vector space of the constructed multi-relational complex network increases correspondingly.

The user-item rating network and user trust relationship network are loaded into a multi-relational complex network by subnet loading operation. The loading process of subnet is as follows: Suppose there are vector composite networks $\Sigma_1 = (G_1, S_1, M_1)$ and $\Sigma_2 = (G_2, S_2, M_2)$, where $G_1 = (V_1, E_1, R_1, F_1)$, $G_2 = (V_2, E_2, R_2, F_2)$, and the set of relation is R', G_1 is the base network. Loading Σ_2 to Σ_1 under load mapping $\Psi: V_1 \times V_2 \to 2^{R'}$ and relation strength mapping $\Psi: (V_1 \times V_2) \times R' \to \cup_{r' \in R'} dom(r')$, which generates a new composite network $\Sigma = (G, S, M)$, where $G = (V, E, R, F)$, R' is called the set of loading relations, $\langle v_h, v_l \rangle \in V_1 \times V_2$ represents the external edge, and v_h, v_l represents the boundary node.

Compared with the traditional complex network, the topological properties of the composite network after loading have changed differently. Under the circumstances, the edges associated with the nodes contain all the relationships of the subnet. Therefore, the mixing weight and mixing degree of the nodes in the composite network are defined as follows:

Mixing weight of node: if $v_h \in V$, the ratio coefficient of relationship strength expressed as $SF = (sf_1, sf_2 \dots sf_k)$, and the mixing weight $k_{v_h}^R$ of node v_h is expressed as:

$$k_{v_h}^R = \sum_{r_i \in R} sf_i \times k_{v_h}^{r_i} \tag{1}$$

where $k_{v_h}^{r_i}$ is the weight of node v_h about the relation r_i.

Mixing degree of node: if $v_h \in V$, the ratio coefficient of relationship strength expressed as $SF = (sf_1, sf_2 \dots sf_k)$, and the mixing degree $\hat{k}_{v_h}^R$ of node v_h is expressed as:

$$\hat{k}_{v_h}^R = \sum_{r_i \in R} sf_i \times \hat{k}_{v_h}^{r_i} \tag{2}$$

where $\hat{k}_{v_h}^{r_i}$ is the weight of node v_h about the relation r_i.

The user-item rating network R_G of Fig. 3 and the user trust relationship network T_G of Fig. 4 are loaded into a composite recommendation network, where the relationship strength parameters sf_1 and sf_2 of relation r_1 and r_2 are set to 1, the edge weight set of subnet R_G is $\{1\}$, the edge weight set of subnet T_G is $\{0.5\}$, and the composite network of sub-network is constructed as shown in Fig. 5:

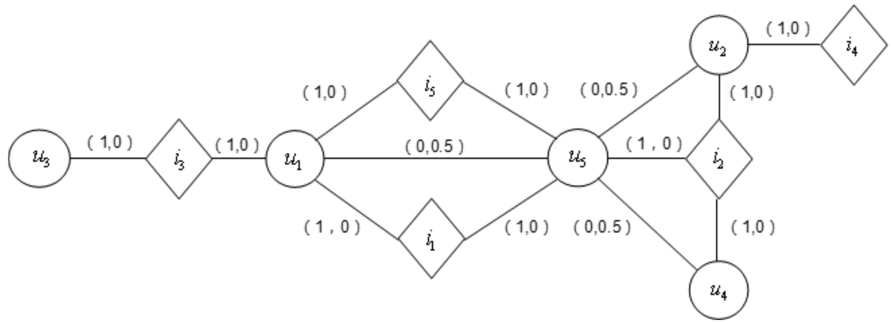

Figure 5: Composited recommendation network

where node u_5 has a degree $\hat{k}_{v_h}^{r_1} = 3$ of relation r_1 and a degree $\hat{k}_{v_h}^{r_2} = 3$ of relation r_2, then the mixing degree of node u_5 is $\hat{k}_{v_h}^{R} = 1 \times \hat{k}_{v_h}^{r_1} + 1 \times \hat{k}_{v_h}^{r_2} = 6$;on the other hand, node u_5 has a weight $k_{v_h}^{r_1} = 1 + 1 + 1 = 3$ of relation r_1 and a weight $k_{v_h}^{r_2} = 0.5 + 0.5 + 0.5 = 1.5$ of relation r_2, then the mixing weight of node u_5 is $k_{v_h}^{R} = 1 \times k_{v_h}^{r_1} + 1 \times k_{v_h}^{r_2} = 4.5$.

4 Community structure detection in multi-relationship social networks

4.1 The process of information propagation on complex networks

Initially, all nodes in the network do not carry information. At the beginning of the first propagation, a node i is set as the source node and given its non-empty information quantity I, and the set of neighbor nodes of the node is set as V^i. At this time, the information in the network is distributed as $X = (0, \dots, I, \dots, 0)$, source nodes transmit information to nodes in V^i by connecting edges. After receiving information, the neighbor nodes will continue to transmit information to their neighbor nodes of the set V^{ii}. However, the node as the receiver does not receive the information from the source node completely, but receives it with a certain probability p. In the second propagation, $V^i = (0, \dots, I_1, \dots, 0, \dots, I_j, \dots, 0, \dots, I_n)$ represents the set of nodes carrying information, and the nodes in the set V^i can either transmit information to neighboring nodes or receive information from their neighboring nodes. Because the receiving node cannot receive all the foreign information, when the network scale is large, it only relies on a single dissemination, and some nodes far from the source node receive less information, which cannot adequately express the impact of the source node on these nodes. Accordingly, the distribution of information carried by all nodes in the network can represent the influence of source nodes on the information dissemination of the whole network only after the process iterates for T times. Thus, the similarity between nodes can be calculated according to the distribution vectors of information of the network when each node acts as the source node.

As a matter of fact, information dissemination can be extended to multi-relational social networks. $R = (r_1, r_2, \dots r_i, \dots, r_n)$ represents the set of relationships existing in the network. Based on the multi-subnet complex network model, a composite network containing n kinds of relationships can be constructed. Similar to single-relationship complex networks, each node in the composite network receives information according to a certain probability p, the difference is that it needs to consider the influence of the interaction of various relationships between nodes and neighbor nodes on probability p.

The information dissemination probability matrix $\widehat{P}$ of composite network is defined as follows:

$$\widehat{P} = \widehat{D}^{-1}\widehat{W} \tag{3}$$

where, $\widehat{D}$ is a diagonal matrix composed of the mixing degree of nodes, $\widehat{W}$ is the adjacent mixing weight matrix, and the elements in the matrix can be expressed as follows:

$$\widehat{W}_{v_h,v_l} = \sum_{(v_h,v_l)\in V} sf_{r_i} \times k'_{r_i} \tag{4}$$

The elements in the adjacent mixing weight matrix are the weighted summation of the strength of the relationships between node v_h and v_l, where sf_{r_i} denotes the proportional coefficient of the strength of relation r_i between node v_h and v_l, and k'_{r_i} denotes the strength of relation r_i.

On the basis of adjacency matrix, the mixing weight of node v_h can also be expressed as:

$$k^R_{v_h} = \sum^n_{v_l\in V} \widehat{W}_{v_h,v_l} \tag{5}$$

In turn, the nodes of the network are selected as the source nodes and given 1 unit of information, and the information quantity of the remaining nodes is 0. Taking node s as an example, assuming that node s is the source node, its carrying information quantity is 1, and the vector of node s is initialized to $X^0_s = (0, 0, \dots, 1, 0)$. The first process is that the node transmits information to its neighbors, which can be expressed as $X^1_s = \widehat{P}X^0_s$ by mathematical formula, and the element in vector X^1_s is the information quantity carried by each node after the first dissemination; in the second process, the nodes have receiving information transmits the information to its neighboring node, and if the neighboring node has information, it can also receive the information from its neighboring node, thus the process can be expressed as $X^2_s = \widehat{P}X^1_s$, and the element in vector X^2_s is the information quantity carried by each node after the second dissemination; Similarly, the information after each dissemination can be expressed as $X^{t+1}_s = \widehat{P}X^t_s$; Before the next information dissemination, assume that $X^{t+1}_s = 1$, which means that the information quantity of the source node is always 1 at the beginning of $t + 1$ process. In this way, the information distribution matrix which represent the influence of nodes can be obtained by using the information distribution vector of each node after T times dissemination.

Taking the composite network in Fig. 5 as an example, its information dissemination probability matrix $\widehat{P}$ is expressed as:

$$P=\begin{vmatrix}4&0&0&0&0&0&0&0&0&0\\0&3&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&0\\0&0&0&2&0&0&0&0&0&0\\0&0&0&0&6&0&0&0&0&0\\0&0&0&0&0&2&0&0&0&0\\0&0&0&0&0&0&3&0&0&0\\0&0&0&0&0&0&0&2&0&0\\0&0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0&0&2\end{vmatrix}^{-1} * \begin{vmatrix}0&0&0&0&0.5&1&0&1&0&1\\0&0&0&0&0.5&0&1&0&1&0\\0&0&0&0&0&0&0&1&0&0\\0&0&0&0&0.5&0&1&0&0&0\\0.5&0.5&0&0.5&0&1&1&0&0&1\\1&0&0&0&1&0&0&0&0&0\\0&1&0&1&1&0&0&0&0&0\\1&0&1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0&0\\1&0&0&0&1&0&0&0&0&0\end{vmatrix}$$

Node u_1, node u_2 and node i_5 are used as source nodes respectively for information dissemination, and the vector of first dissemination is obtained as follows:

X_1=[0, 0, 0, 0, 0.13, 0.25, 0, 0.25, 0, 0.25]

X_2=[0, 0, 0, 0, 0.17, 0, 0.33, 0, 0.33, 0]

X_3=[0.5, 0, 0, 0, 0.5, 0, 0, 0, 0, 0]

Further, the vector of first dissemination is obtained as follows:

X_1=[0.39, 0.01, 0.13, 0.01, 0.38, 0.27, 0.02, 0.25, 0, 0.27]

X_2=[0.01, 0.46, 0, 0.13, 0.28, 0.03, 0.36, 0, 0.33, 0.03]

X_3=[0.54, 0.04, 0, 0.04, 0.56, 0.21, 0.08, 0.13,0,0.21]

4.2 Community detection algorithm for multi-relationship social networks based on information propagation

After the nodes of network are represented by vectors, their data structure is suitable to be processed by clustering algorithm, and the similarity between nodes is calculated according to the information dissemination, thus the mature data mining clustering algorithm is applied to the community detection for multi-relational social networks.

In previous studies, CSDM algorithm [15] was proposed based on classical K-means clustering algorithm to discover the community structure of multi-relational social networks. The biggest disadvantage of K-means algorithm is that it is sensitive to the selection of initial clustering centers, since the initial value of the cluster center will directly affect the final clustering result.

Generally, the importance of nodes in complex networks is called centrality, which is the basic index of describing the importance of nodes. Degree centrality considers that the influence of nodes is related to the number of neighbor nodes, which means that the more neighbor nodes, the greater the influence of the node in the network. Assuming that the path length of node v_h and v_l with respect to relation r_i is $\hat{l}^{r_i}_{v_h v_l}$, and the centrality of node v_h with respect to relation r_i is defined as:

$$\hat{c}^{r_i}_{v_h} = \frac{\min_{v_g \in V} \sum_{v_l \in V} \hat{l}^{r_i}_{v_g v_l}}{\sum_{v_l \in V} \hat{l}^{r_i}_{v_h v_l}} \tag{5}$$

Where the molecular part represents the minimum sum of the path lengths of the relationship r_i with the remaining nodes, and the denominator part represents the sum of the path lengths of the node v_h and other nodes with respect to the relationship r_i. Specifically, the closer the value of centrality $\hat{c}_{v_h v_l}^{r_i}$ is to 1, the smaller the sum of the path length of the node and the remaining nodes is, the stronger the centrality will be; conversely, the closer the value of centrality is to 0, the weaker the centrality will be.

Further, the CSDM algorithm chooses the nodes with larger centrality as the initial clustering centers, and then divides the multi-relational social networks into communities according to classical *K*-means clustering algorithm.

5 Collaborative filtering recommendation algorithm based on community structure detection

Through information dissemination, it solves the problem that the social network composed of users of recommendation system is difficult to calculate similarity due to multi-relationship, and the information quantity carried by each member after dissemination reflects the similarity degree of different nodes in the network. By dividing these users with high similarity into a community, members of the community will share information such as rating preferences and interest preferences. On the one hand, the social preference vector can be obtained from the trust relationship of users contained in the community structure, which represents the user's social preference and can be measured by social similarity; on the other hand, users' interest preferences can be obtained from the rating information contained in the community structure, and the similarity of interest preferences can be determined by extracting interest preference vectors. Finally, the social similarity and interest preference similarity are synthesized in a certain proportion, and a new comprehensive similarity index, recommendation degree, is obtained, which is used to rank and recommend commodities.

5.1 User social similarity and rating similarity

Assuming that there is a social relationship between user u and user v, and the interest between u and v is closer, thus their preference similarity is higher than that of strangers [Wang and Ma (2016)]. The kind of social relationship with promoting recommendation effect can be expressed as trust relationship between users. In this way, users in the same community are mapped to a new user trust network, and they are mapped to vectors by information dissemination method, then Pearson similarity is used to measure the social similarity between users u and v, the formula is as follows:

$$sim_{social}(u,v) = \frac{\sum_l (r_{ui}-\bar{r}_u)(r_{vi}-\bar{r}_v)}{\sqrt{\sum_l (r_{ui}-\bar{r}_u)^2}\sqrt{\sum_l (r_{vi}-\bar{r}_v)^2}} \tag{6}$$

In fact, users' common rating items and rating preference also reflect users' interests to a certain extent. In addition to Pearson correlation to measure the user's rating preference, the user preference degree [11] is defined, and the more consistent the rating preference is, the closer the user's interest preference is, the formula is as follows:

$$P(u,v) = \left(\frac{|I_u \cap I_v|^{(\geq R)}}{|I_u|^{(\geq R)}} + \frac{|I_u \cap I_v|^{(<R)}}{|I_u|^{(<R)}}\right) \times \left(\frac{|I_u \cap I_v|^{(\geq R)}}{|I_v|^{(\geq R)}} + \frac{|I_u \cap I_v|^{(<R)}}{|I_v|^{(<R)}}\right) \tag{7}$$

where the threshold value of R is consistent with the threshold value when constructing user-item rating network. In Section 3.1, the user's rating below the threshold R indicates that the user does not have a preference relationship for the commodity. While, in this case, it is used to indicate that the user's rating is positive or negative. If the rating is greater than the threshold R, it means that the rating is positive; $\mathrm{I_u}$ and $\mathrm{I_v}$ represent the set of commodities scored by user u and v respectively. When user preference is integrated into the process of calculating the rating similarity, the user's rating of commodities in the community constitutes a new rating vector, $sim(u,v)$ is the Pearson correlation between the rating vector of u and the rating vector of v. Thus, user rating similarity based on rating preference is obtained:

$$sim_{rating}(u,v) = P(u,v) \times sin(u,v) \tag{8}$$

In order to improve the reliability of user similarity, the following formulas are used to integrate user social similarity and user rating similarity:

$$\mathrm{sim}(u,v) = (1-\alpha) * sim_{social} + \alpha * sim_{rating} \tag{9}$$

where parameter α controls the weight of the two similarities in the recommendation process, $\mathrm{sim_{social}}$ is the social similarity between user u and v, and $\mathrm{sim_{rating}}$ is the rating similarity. When α is greater than 0.5, rating similarity accounts for a larger proportion of the final similarity, while α is less than 0.5, social similarity is the main factor.

5.2 Prediction of rating

The recommendation system mainly relies on the prediction of the items that are not scored by the recommended users, and finally the top-n with the highest rating will be recommended. It is not necessary to use all users in the community when predicting. Usually, k sets of nearest neighbors with the highest similarity of current users are selected. In the process of selecting the nearest neighbor users, if the value of k is too large, additional noise will be brought to the prediction when the neighbors with low similarity are added; if the value of k is too small, it will lead to deviation due to insufficient nearest neighbors used to predict. The prediction rating of user u about commodity i is calculated by the following formula:

$$r_{ui} = \bar{r}_u + \frac{\sum_{v \in S(u)}(r_{vi} - \bar{r}_v) \times sim(u,v)}{\sum_{v \in S(u)} sim(u,v)} \tag{10}$$

where $\mathrm{S_u}$ is the set of k users with the highest similarity to user u in the community, $\mathrm{r_{vi}}$ is the rating of user v about commodity i in $\mathrm{S_u}$, and $\mathrm{\bar{r}_v}$ is the average value of all commodities scored by user v.

6 Experiment and result analysis

In the process of experiment, the effectiveness and recommendation quality of the proposed algorithm are proved by comparing with different classical recommendation algorithms, and the influence of parameter setting on the recommendation results is analyzed as well.

6.1 Experimental environment and dataset

In order to analyze big dataset, a Hadoop cluster platform composed of 5 nodes is built, and the specific configuration of each node machine is Intel (R) Core (TM) i7-7700 CPU@3.60 GHz, 8 GB memory, 64-bit operating system.

Epinions dataset [Lauw, Lim and Wang (2012)] commonly used in recommendation system is used in the experiment. The dataset contains users' rating information and social information. There are 40163 users and 139738 commodities in the dataset, and the rating records reach 664824 items. The user's rating of commodities is expressed by 1 to 5. If there is trust relationship between users, it is expressed by 1; if there is no trust relationship, it is marked as 0; and other statistical information in the dataset is shown in Tab. 1:

Table 1: The statistics of Epinions dataset

Name	Minimum rating quantity	Maximum rating quantity	Average rating quantity
User	1	1022	16.55
Item	1	2018	4.76

In order to use the algorithm proposed for rating prediction, it is necessary to determine the relationship coefficients and strength between rating relationship and trust relationship. For this reason, users with number 2 and number 10004 are chosen as seed nodes. After eliminating some obvious outliers, 150 sample data were selected as the sample set of linear regression analysis data for calculating the proportional coefficients of the two relationships. The linear regression model was constructed for rating relationship r_1 and trust relationship r_2 in multi-relational social networks as follows:

$$y = a_0 + sf_1 x_1 + sf_2 x_2 \tag{11}$$

where a_0, sf_1 and sf_2 are regression coefficients, the least squares method is used to estimate the parameters and the equations are as follows:

$$\begin{cases} \sum y = \mathrm{sf}_1 \sum x_1 + \mathrm{sf}_2 \sum x_2 \\ \sum x_1 y = \mathrm{sf}_1 \sum x_1^2 + \mathrm{sf}_2 \sum x_1 x_2 \\ \sum x_2 y = \mathrm{sf}_1 \sum x_1 x_2 + \mathrm{sf}_2 \sum x_2^2 \end{cases} \tag{12}$$

The average value of the parameter estimation is obtained as follows: $sf_1 = 2.0, sf_2 = 2.8$.

Therefore, when dividing the community of the multi-relationship social network constructed in the experiment, assumed that the ratio coefficient of the two relationships is $SF = (2.0, 2.8)$, the strength of the relationship is 1, the parameter of the K-means clustering algorithm is $k = 5$, and the community partition is shown in Fig. 6.

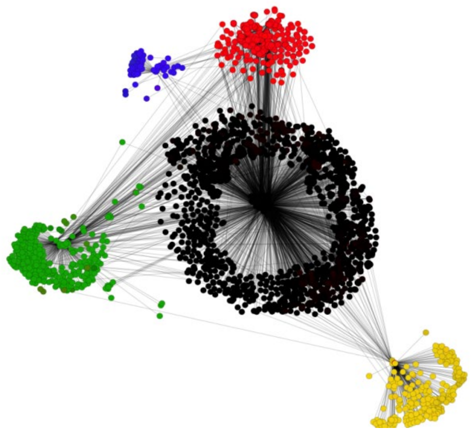

Figure 6: Result of community detection

6.2 Analysis and comparison of experimental results

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are selected as evaluation indicators to evaluate the accuracy of recommendation results, which evaluate the accuracy of recommendation algorithm by comparing the errors between the real rating and the predicted rating. The smaller the value, the better the recommended effect is. The specific definitions are as follows:

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\left|r_{i,j} - \grave{r_{i,j}}\right| \tag{13}$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i,j}\left(r_{i,j} - \grave{r_{i,j}}\right)^2} \tag{14}$$

where, $r_{i,j}$ denotes the real rating of user i for commodity j, $\grave{r_{i,j}}$ denotes the predictive rating of user i for commodity j, and n is the number of ratings in the test set.

The selection of the nearest neighbor numbers k will affect the experimental results, and too small or too large of the nearest neighbor set will increase the uncertainty of prediction. In the experiment, k of the proposed algorithm are set to be 10 and 30, respectively. As shown in Tab. 2, it is clear that MAE and RMSE of k at 30 are significantly lower than 10. Accordingly, the nearest neighbor numbers $k = 30$ is set in the following comparison experiments of kinds of recommendation algorithms.

Table 2: Influence of different neighbor numbers on algorithm

Neighbor numbers	MAE	RMSE
$k = 10$	0.62	0.82
$k = 30$	0.44	0.59

In the experiment, the proposed algorithm (CDRec) is compared with three commonly used social recommendation algorithms TDSRec [Guo, Wang and Hou (2018)], Social MF [Forsati, Mahdavi and Shamsfard (2014)] and TDRec [Park, Kim and Oh (2016)], and the results are shown in Fig. 7.

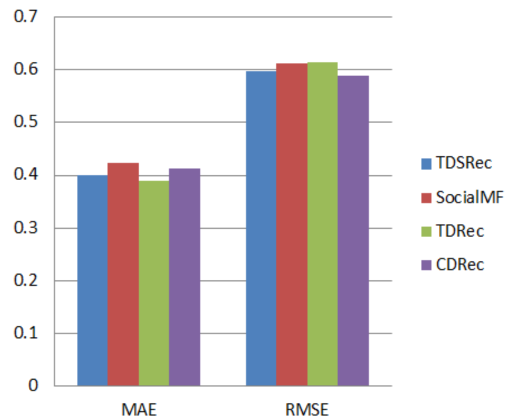

Figure 7: The results comparison of four recommendation algorithms

From Fig. 7, it is can be seen that the proposed algorithm is similar to other social recommendation algorithms on MAE and slightly better than other algorithms on RMSE, which means that the proposed algorithm can predict ratings of the recommended items more accurately.

In view of the datasets vary in scale, the execution comparison of four social recommendation algorithms is shown in Fig. 8.

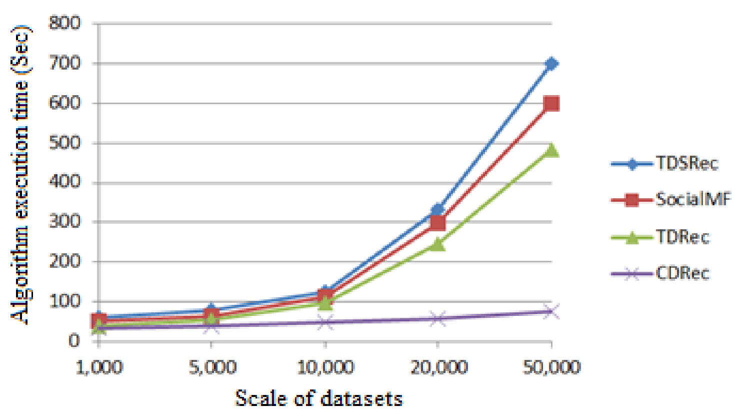

Figure 8: The execution comparison of four recommendation algorithms

As can be seen from Fig. 8, with increase of the scale of dataset, the execution of the proposed algorithm is significantly less than that of the other three social recommendation algorithms.

In conclusion, the proposed algorithm can improve the recommendation efficiency on the basis of guaranteeing the accuracy of recommendation when dealing with big sample datasets.

7 Conclusion

For recommendation system in big data, the efficiency of recommendation al-gorithm is very important. In this paper, a collaborative filtering algorithm based on multi-subnet complex network model is proposed to construct multi-relationship complex network. By using community structure detection, similar users can be partitioned into a community, which greatly reduces the computa-tional complexity of recommendation algorithm. Experiments on real data sets show that the proposed algorithm can improve the efficiency of recommendation algorithm on the basis of guaranteeing the accuracy of recommendation.

Acknowledgement: This work is supported by the Humanity and Social Science Youth foundation of Ministry of Education of China under Grant No. 15YJC860001. This research is also supported by Shandong Provincial Natural Science Foundation, China under Grant No. ZR2017MG011 and Shandong Social Science Research Project under Grant Nos. 17CHLJ16 and 18BJYJ04. This work is also supported by the Grant of Fujian Middle-aged and Young Teachers' Educational and Scientific Research Projects in 2018, Grant No. JZ180469.

References

Chang, W. L.; Diaz, A. N.; Hung, P. C. K. (2015): Estimating trust value: a social network perspective. *Information Systems Frontiers*, vol. 17, no. 6, pp. 1381-1400.

Dong, C. L.; Ke, X. S. (2018): Study on collaborative filtering algorithm based on user interest change and comment. *Computer Science*, vol. 45, no. 3, pp. 213-217.

Forsati, R.; Mahdavi, M.; Shamsfard, M. (2014): Matrix factorization with explicit trust and distrust side information for improved social recommendation. *ACM Transactions on Information Systems*, vol. 32, no. 4, pp. 17-27.

Guo, L. J.; Li, N.; Zhao, X. W. (2016): Collaborative filtering recommendation algorithm incorporating social network information. *Pattern Recognition and Artificial Intelligence*, vol. 29, no. 3, pp. 281-288.

Guo, N. N.; Wang, B. L.; Hou, Y. H. (2018): Collaborative filtering recommendation algorithm based on characteristics of social network. *Journal of Frontiers of Computer Science and Technology*, vol. 16, no. 2, pp. 208-217.

Jiang, M. M.; Sun, G. X.; Bin, S. (2018): Community detection algorithm in multiple relationships online social network. *Journal of Frontiers of Computer Science and Technology*, vol. 11, no. 7, pp. 1-13.

Jiang, S.; Qian, X.; Shen, J. (2015): Author topic model-based collaborative filtering for personalized POI recommendations. *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 907-918.

Lauw, H. W.; Lim, E. P.; Wang, K. (2012): Quality and leniency in online collaborative rating systems. *ACM Transactions on the Web*, vol. 6, no. 1, pp. 1-27.

Li, Y.; Li, J.; Chen, J.; Lu, M. (2018): Seed selection for data offloading based on social and interest graphs. *Computers, Materials & Continua,* vol. 57, no. 3, pp. 571-587.

Liu, H. X. (2016): A survey of collaborative filtering technique in recommendation system. *Information Security and Technology*, vol. 7, no. 3, pp. 24-26.

Margaris, D.; Vassilakis, C.; Georgiadis, P. (2018): Query personalization using social network information and collaborative filtering techniques. *Future Generation Computer Systems*, vol. 78, no. 7, pp. 440-450.

Pan, J. C.; Zhang, X. M.; Wang, X. (2016): Improved singular value decomposition recommender algorithm based on user reliability. *Journal of Chinese Computer Systems*, vol. 37, no. 10, pp. 2171-2176.

Park, C.; Kim, D.; Oh, J. (2016): Improving top-k recommendation with truster and trustee relationship in user trust network. *Information Sciences*, vol. 374, no. 3, pp. 100-114.

Roh, T. H.; Oh, K. J.; Han, I. (2003): The collaborative filtering recommendation based on SOM cluster-indexing CBR. *Expert Systems with Applications*, vol. 25, no. 3, pp. 413-423.

Santhini, M.; Balamurugan, M.; Govindaraj, M. (2015): Collaborative filtering approach for big data applications based on clustering. *International Journal of Mathematics Computer Science and Information Technology*, vol. 2, no. 1, pp. 202-208.

Sun, G. X.; Bin, S. (2018): A new opinion leaders detecting algorithm in multi-relationship online social networks. *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4295-4307.

Sun, G. X.; Bin, S. (2017): Router-level internet topology evolution model based on multi-subnet composited complex network model. *Journal of Internet Technology*, vol. 18, no. 6, pp. 1275-1283.

Wang, M.; Ma, J. (2016): A novel recommendation approach based on users' weighted trust relations and the rating similarities. *Soft Computing*, vol. 20, no. 10, pp. 3981-3990.

Wang, Q.; Li, W.; Zhang, X. (2016): Academic paper recommendation based on community detection in citation-collaboration networks. *Lecture Notes in Computer Science*, vol. 9932, no. 1, pp. 124-136.

Zhang, H. X.; Zhen, L.; Zhang, C. T. (2018): An improved collaborative filtering recommendation algorithm with weighted heterogeneous information. *Journal of University of Electronic Science & Technology of China*, vol. 47, no. 1, pp. 112-116.

Zhang, H. Y.; Ding, F.; Jiang, L. H. (2005): A collaborative filtering recommendation method based on fuzzy clustering. *Computer Simulation*, vol. 22, no. 8, pp. 144-147.

Zhang, Y.; Li, S. Y.; Tang, Y. (2017): Research on heterogeneous knowledge fusion algorithm under big data environment. *Computer Technology and Development*, vol. 27, no. 9, pp. 12-16.

Zhao, Q. Q. (2013): SPCF: a memory based collaborative filtering algorithm via propagation. *Chinese Journal of Computers*, vol. 36, no. 3, pp. 671-676.