Research Article

An Improved Collaborative Filtering Recommendation Algorithm and Recommendation Strategy

Xiaofeng Li [ID][1] **and Dong Li**[2]

[1]*Department of Information Engineering, Heilongjiang International University, Harbin 150025, China*
[2]*School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China*

Correspondence should be addressed to Xiaofeng Li; lipf2012@126.com

The e-commerce recommendation system mainly includes content recommendation technology, collaborative filtering recommendation technology, and hybrid recommendation technology. The collaborative filtering recommendation technology is a successful application of personalized recommendation technology. However, due to the sparse data and cold start problems of the collaborative recommendation technology and the continuous expansion of data scale in e-commerce, the e-commerce recommendation system also faces many challenges. This paper has conducted useful exploration and research on the collaborative recommendation technology. Firstly, this paper proposed an improved collaborative filtering algorithm. Secondly, the community detection algorithm is investigated, and two overlapping community detection algorithms based on the central node and k-based faction are proposed, which effectively mine the community in the network. Finally, we select a part of user communities from the user network projected by the user-item network as the candidate neighboring user set for the target user, thereby reducing calculation time and increasing recommendation speed and accuracy of the recommendation system. This paper has a perfect combination of social network technology and collaborative filtering technology, which can greatly increase recommendation system performance. This paper used the MovieLens dataset to test two performance indexes which include MAE and RMSE. The experimental results show that the improved collaborative filtering algorithm is superior to other two collaborative recommendation algorithms for MAE and RMSE performance.

1. Introduction

As the amount of information continues to increase, the rapid development of Internet technology has brought us into the era of information explosion. The simultaneous appearance of massive information makes it difficult for users to find the parts they are interested in; on the other hand, it also makes a large number of people's information to be "dark information" in the network, which cannot be obtained by ordinary users. Therefore, people need to spend a lot of time browsing and finding information of interest. People use search engines to retrieve information, such as traditional Google, Yahoo, and Baidu. Although information retrieval technology can meet people's needs to a certain extent, due to its universal characteristics, it still cannot meet user requests of different backgrounds, different purposes, and different periods. Personalized recommendation service technology

came into being. It can provide different services for different users to meet the need for personalized services [1, 2]. There are mainly some problems in the current recommended technologies, like the more common collaborative recommendation technology, such as user-based and item-based collaborative filtering algorithms. Due to the sparsity of data, new users, and new items, the performance and accuracy of the recommendation system are limited. In addition, some online recommendation systems, real-time recommendation, are not guaranteed due to large-scale processing of data. Therefore, it is still a hot topic to build a high accuracy and high scalability recommendation algorithm. As an important branch of data mining, social network analysis has received more and more attention [3]. Social network analysis is a kind of link analysis technology, which uses social networks as a research object to analyze the structure and behavior of social networks. Since there is a large amount of information in the

social network that can be used for analysis and mining, social network analysis has been introduced into various application fields. The existing P2P trust technology has actually adopted the social network analysis method. It has been shown that the personalized recommendation based on the social network can solve the problem of data sparseness, new users, and new items in the traditional personalized recommendation. Based on the research of existing recommendation technologies, this paper proposes to introduce social network analysis technology into the recommendation system to realize a high efficiency and high scalability recommendation system to solve the problems of new users and new items [4].

The collaborative filtering recommendation technique has both practical values and shortcomings. From the perspective of recommendation system implementation, the memory-based collaborative recommendation is to load the user and product information into the internal memory and carry out real-time recommendations through the calculation [5–7]. It is characterized by simple implementation and no training overhead [8, 9]. However, the large amount of information will lead to excessive system overhead and slow recommendation speed. Although the model-based recommendation is faster, the system will retain the data to obtain the new model and produce large overhead when adding new users, new items, or new scoring [10, 11]. In recent years, there have been many improved recommendation algorithms based on collaborative filtering to solve the problems. For example, Rashid et al. [12] put forward the collaborative filtering algorithm based on the clustering model; Xue et al. proposed to adopt the clustering smooth collaborative filtering algorithm. Therefore, how to create an accurate, fast, and effective high-scalable recommendation system has become the trend of research. This paper combines the social network technology with the collaborative filtering recommendation technique, applies the community detection idea to the collaborative recommendation, and adopts the scoring pretreatment mechanism during the collaborative recommendation to prevent the data scarcity [13]. Afterwards, this paper introduces the improved community detection algorithm and organically combined the community detection algorithm with the collaborative recommendation algorithm and studied the collaborative filtering recommendation algorithms based on the community detection.

2. Improved Community Detection Algorithm

2.1. Overlapping Community Detection Algorithm Based on Central Nodes

2.1.1. Basic Idea. Taking the "central node" as the initial community C, this paper adds the neighbor node with the largest contribution to the community C to this community, and a community will be built when the global contribution reaches the maximum; if there are more nodes with larger contribution to the community, these nodes will be added to these communities. After extracting the community C, the nodes and edges of community C are not deleted from the network, in order to mine the overlapping nodes between communities.

2.1.2. Basic Concept. To illustrate the algorithm proposed in this section, first define the following.

(1) Local contribution q [14]:

$$q = \frac{L_{in}}{L_{in} + L_{out}}, \tag{1}$$

L_{in}: if it is not entitled to map, it represents the number of internal sides of the community; if it is a right to map, it is the sum of the weights of all the edges within the community. L_{out}: if it is not entitled to map, it represents the number of external sides of the community; if it is the right to map, it is the sum of the weights of all the sides of the community and the outside. q: the greater the q, the greater the contribution to the community, and vice versa.

(2) Global contribution degree Q: this represents the current maximum contribution in the community detection process. This indicator is used to determine whether the community structure achieves the best state.

(3) Overlap degree: the overlap of C_i and C_j in the community is defined as S:

$$S = \frac{\left| C_i \cap C_j \right|}{\left| C_i \cup C_j \right|}. \tag{2}$$

2.1.3. Algorithm Flow. Algorithm is divided into two stages. The first stage is to dig the community in the network, and the second stage is to adjust the mining community.

(1) Community mining

Step 1: to calculate the degree of each node in the network, select the largest node i as the initial community C_i, and the node i to do the tag; initial global contribution $Q = 0$.

Step 2: find all the nodes connected to the community C_i in the network, and put them into the neighbor node set U.

Step 3: for each node j in the U, according to equation (1) to calculate the contribution of node j to the community q_j. If the offer J degrees with the largest node degree $q_{max} = \max\{q_j\} \geq Q$, then add node j to the community C_i. Return to step 2 to continue; otherwise, turn to step 4.

Step 4: global contribution of Q has reached a maximum value. Get community C_i.

Step 5: if the network has no unlabeled nodes, all community networks have been detected, the end of the process; otherwise, the node never labeled select nodes as the initial community of new C_i. Then, return to step 2 to continue.

(2) Community adjustment: in fact, some overlapping nodes belong to multiple communities, which have overlapping nodes between communities, community C_i and C_j; when the overlap threshold reached T (0.7), C_i and C_j are closely related, so the community

C_i and C_j merge. The specific process of community adjustment is as follows:

Step 1: to calculate the degree of overlap S between C_i and C_j in any two communities

Step 2: C_i and C_j are merged into a community when the overlap degree of S is greater than the threshold value of T

Step 3: if any of the two communities overlap coefficient is less than the threshold T, adjust the end; otherwise, return to step 1 continue to adjust

(3) Experimental analysis and comparison: we select three classic network datasets:

(1) 34-node Zachary's Karate Club network [15]
(2) 115-node American College Football network [16]
(3) 62-node Dolphins network [17]

Experiment 1, Experiment 2, and Experiment 3, respectively, are testing in the datasets (1), (2), and (3) on the center node based on the overlapping community detection algorithm.

Experiment 1: the dataset Zachary Karate Club network contains 34 nodes. The real network contains two communities.

Through the community dividing according to the overlapping the community detection algorithm based on central nodes, the original network is divided into four communities, where nodes 9, 10, and 31 are overlapping nodes. There is no significant difference between the final community results dividing according to the algorithm and two communities of the real network. Two communities in the funnel network are resegmented. The node sets (5, 6, 7, 11, and 17) and (25, 26, 29, and 32) are extracted as separate communities, and the four communities in Figure 1 are built. In fact, the node set (5, 6, 7, 11, and 17) is closely connected and can be extracted as a separate community; the node set (25, 26, 29, and 32) is less connected to other nodes in the original community and can also be used as a separate community. Moreover, nodes 9, 10, and 31 are boundary nodes of two communities and have an equivalent contribution to two communities, so they can belong to two communities simultaneously. It is shown in Figure 1.

Experiment 2: actually, American College Football network has 115 nodes, and the real network contains 12 independent communities. This paper takes advantage of the overlapping community based on central nodes, mines the American College Football network through the detection algorithm, and mines 10 communities in total. It is shown in Figure 2.

The vast majority of 10 communities dividing according to the overlapping community detection algorithm based on central nodes are consistent with the American College Football network, of which seven communities have exactly the same nodes, and the node division accuracy rate is up to 85% or higher. In the actual network, one community has fewer internal edges than the one between communities. Therefore, when using the proposed algorithm, this paper shall distribute all nodes within the community to other communities. The red nodes (60, 43, 91, 5, 94, 15, 59, and 98) in Figure 2 represent overlapping nodes between two communities. Through careful analysis, it is found that those nodes have larger contributions to the communities they belong to. Therefore, the overlapping nodes are consistent with the actual situation.

Experiment 3: the dataset Dolphins network is the categoric dolphin network. The actual network is divided into two communities. The community is divided by taking advantage of the proposed algorithm, and four communities are obtained. It is shown in Figure 3.

The overlapping community detection algorithm based on central nodes divides the Dolphins network into two communities, where nodes 8 and 40 are overlapping nodes. Compared with two communities in the actual network, except the overlapping nodes, the results of the other nonnodes are identical. The nodes 8 and 40 have the same number of connections with respect to communications and have the same contribution to the two communities, so it is natural for them to belong to two communities as overlapping nodes.

2.2. Overlapping Community Detection Algorithm Based on k-Faction

2.2.1. Basic Idea.
In a sense, the community can be regarded as a set of interconnected "small fully coupled networks". These "fully coupled network" is called the "faction", where k-faction indicates the number of nodes in the fully coupled network is k.

This paper introduces the overlapping community detection algorithm based on k-faction. The algorithm first detects all the k-faction from the network as the initial community. Afterwards, this paper merges the initial community, yet there are still some nodes not adding to any faction. By adding these nodes to the most closely connected community, this paper will get all communities on the network. To make the community division results more accurate, the algorithm finally further optimizes the community, and the optimization criteria are to make the modularity achieve the maximum.

2.2.2. Basic Concept

(1) Faction: the complete graph of the network
(2) The largest faction: the network does not belong to any other faction
(3) Overlap: it is defined as the number of nodes in the community of two/two communities with a small number of nodes in the community
(4) Connectivity between communities: it is defined as the number of connected sides of the two
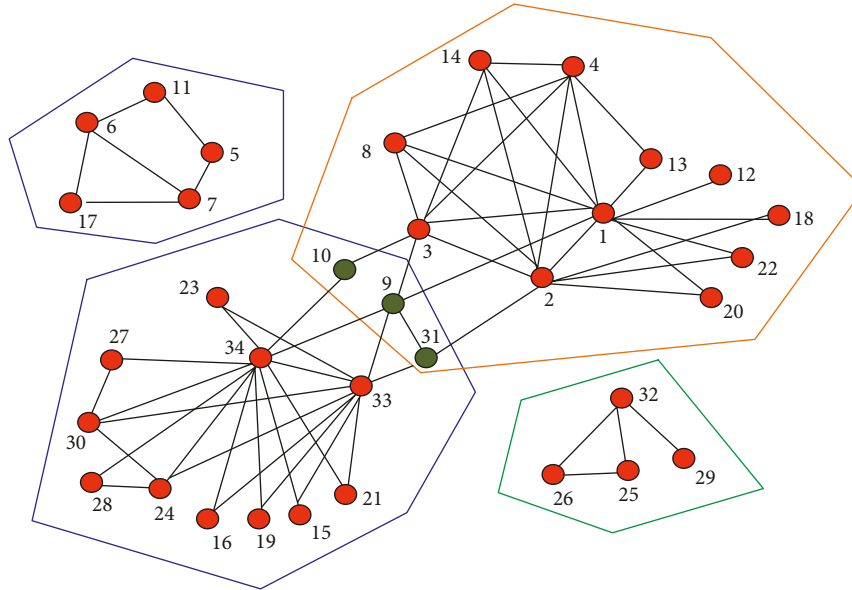
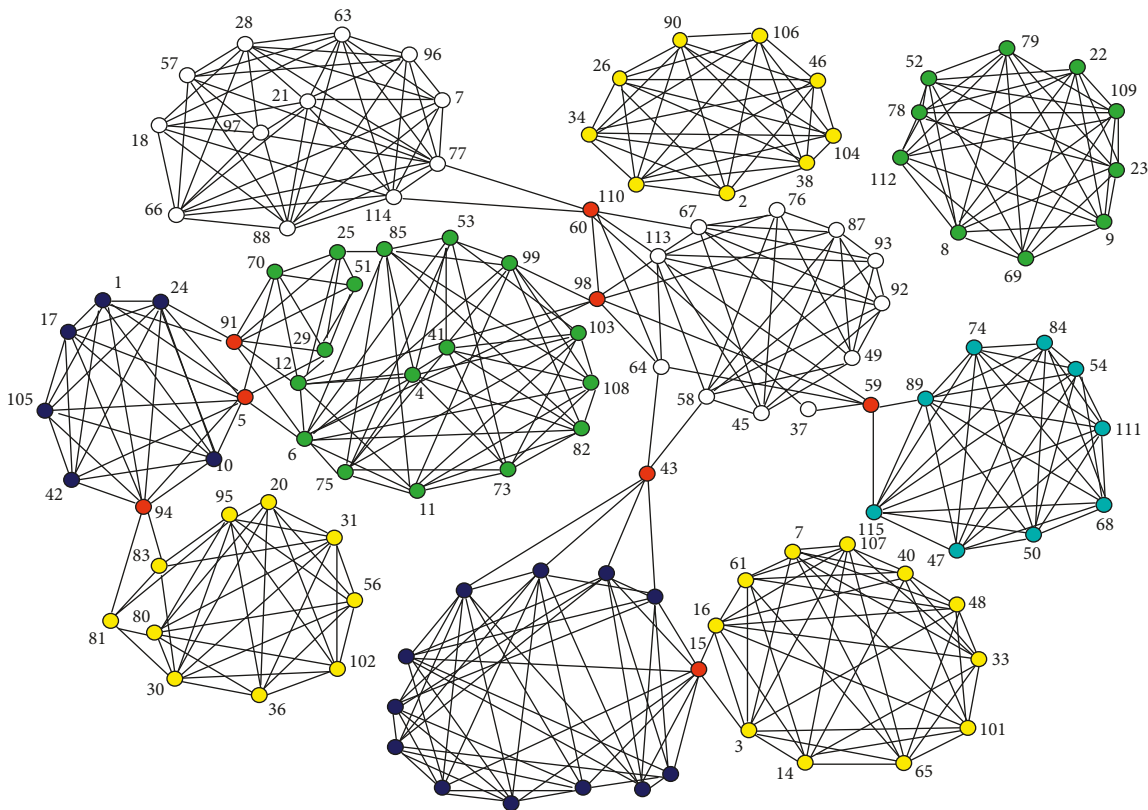FIGURE 1: Overlapping community detection algorithm based on the center node.

FIGURE 2: American College Football network divided into 10 communities.

communities/the total number of connections within the two communities

(5) The point of contact with the community is closely related to M: the number of vertices and edges of the community and the number of vertices in the community

(6) Module Degree Q_c

$$Q_c = \frac{1}{L} \sum_i \sum_{vw} a_{vc_i} a_{wc_i} \left(A_{vw} - \frac{k_v k_w}{L} \right). \tag{3}$$

2.2.3. Algorithm Flow. The algorithm is divided into four processes. First process: from the network to find all the 2 factions more than the largest faction and these
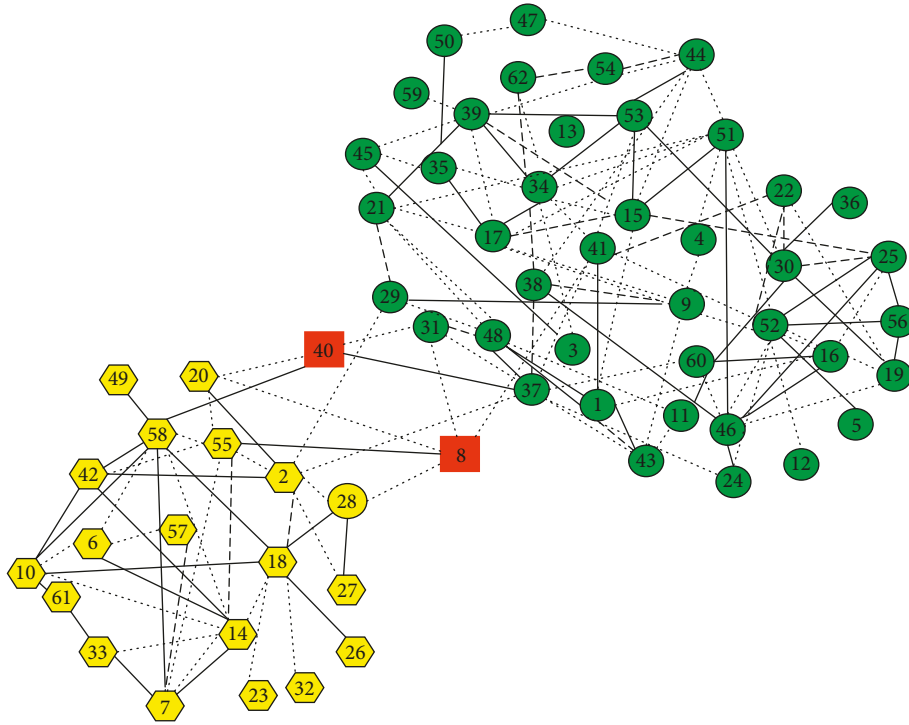
FIGURE 3: Dolphins network divided into 2 communities.

factions as the initial community; second process: according to the community between the point and edge overlap to merge community; third process: join the nodes to join the most close community, community expansion; fourth process: tuning of the second phase of the community.

(1) Finding the largest faction: Bron–Kerbosch algorithm is used to find the maximum clique in the network, and the maximum number of nodes in the network is more than 3 as the initial community.

(2) Merger community: because there are a number of overlapping nodes in all of the biggest factions, these factions can be combined into a community when the overlap is over a threshold value of T; in addition, if the connection between the two communities is closer than the threshold value of CONN, the two communities should also be merged into the same community. The combined community consists of two steps: the first step is based on the point of the merger; the second step is based on the combination of edge.

(3) Experimental analysis and comparison: we select three classic network datasets:

(1) 34-node Zachary's Karate Club network

(2) 115-node American College Football network

(3) 62-node Dolphins network

Experiment 4, Experiment 5, and Experiment 6, respectively, are testing in the datasets (1), (2), and (3) on the center node based on the overlapping community detection

algorithm. The effect of the algorithm is investigated. Experimental parameters are setting CONN = 0.5 and T = 0.65.

Experiment 4: the classic dataset Zachary Karate Club network contains 34 nodes. By using the overlapping community detection algorithm based on k faction, the algorithm is divided into two communities, as shown in Figure 4.

From Figure 4, we can learn that the proposed algorithm is almost identical to the original network, except for the node 10. In fact, 10 nodes and two community connection numbers, as overlapping nodes, were added in two communities in accordance with the actual situation.

Experiment 5: American College Football network is based on the overlapping community detection algorithm based on k-faction. The network is divided into the community, and the results are shown in Figure 5. From Figure 5, we can see that a total of 13 communities, and the original network compared to a total of 87% of the nodes are correctly divided compared to the original network, of which 7 communities are exactly the same, and there are 4 almost identical. In all communities, there is not any overlapping node.

Experiment 6: dataset Dolphins network is a classic dolphin network. The actual network is divided into two communities. Use this section to carry out community division, and get four communities, as shown in Figure 6.

The overlapping community detection algorithm based on the k-faction is divided into four communities,
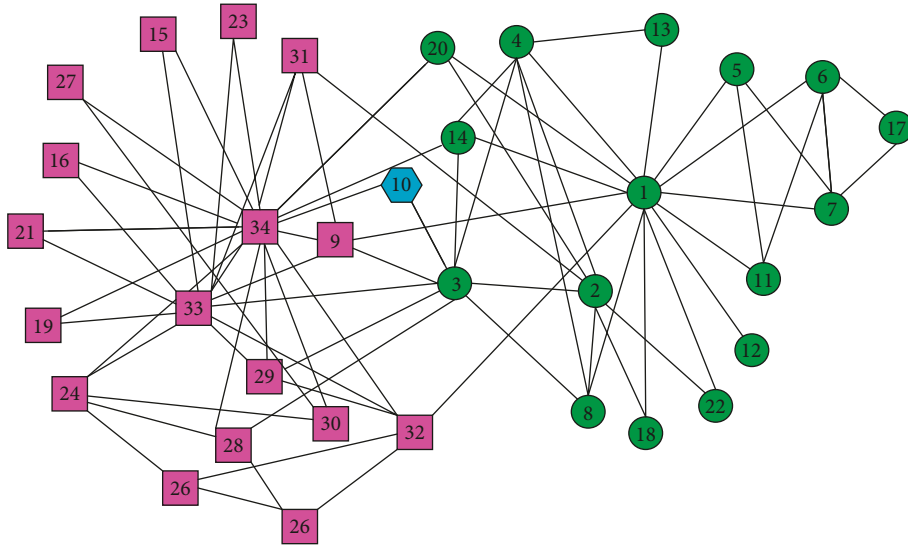
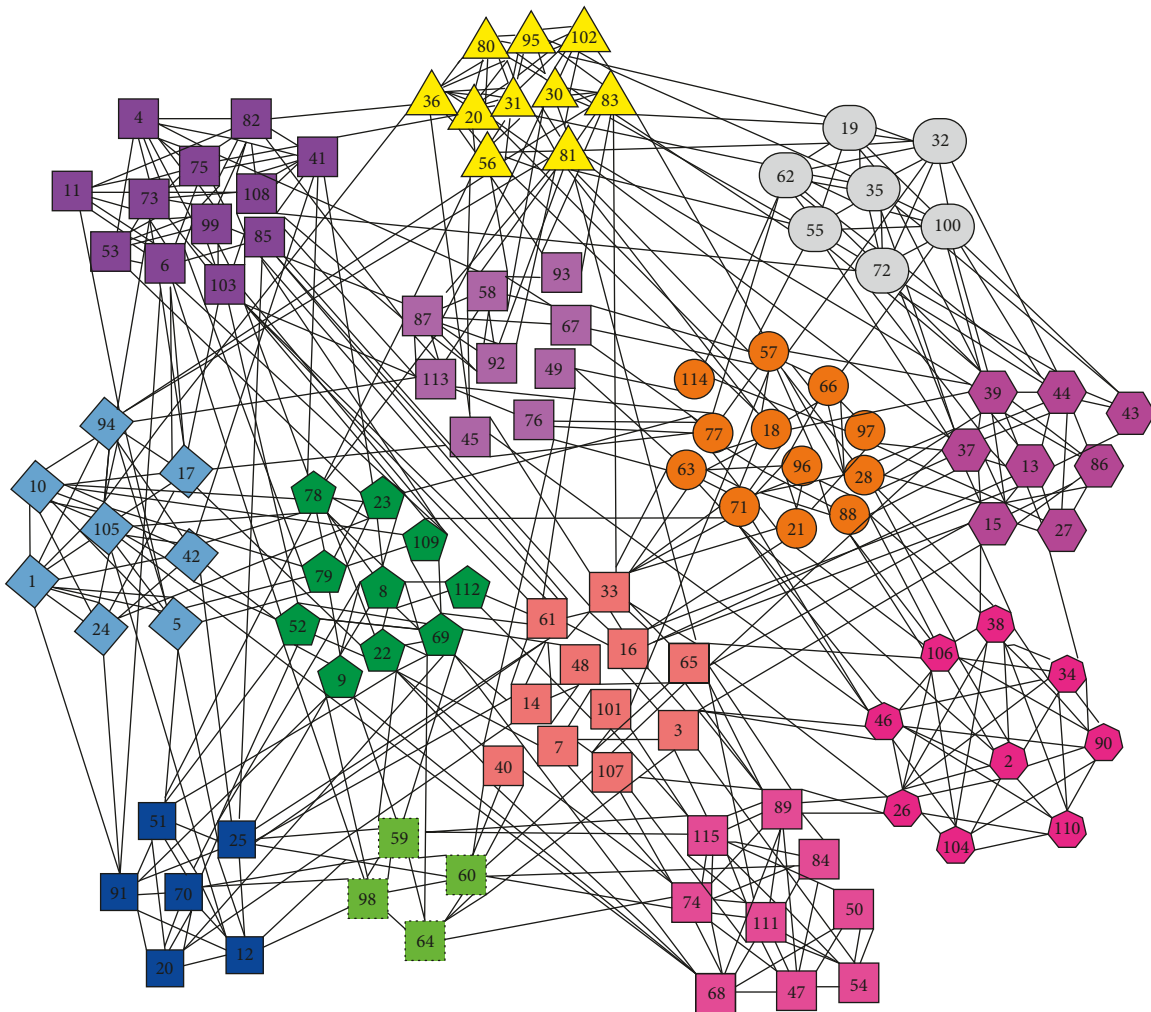FIGURE 4: Community based on the overlapping community detection algorithm based on *k*-faction.

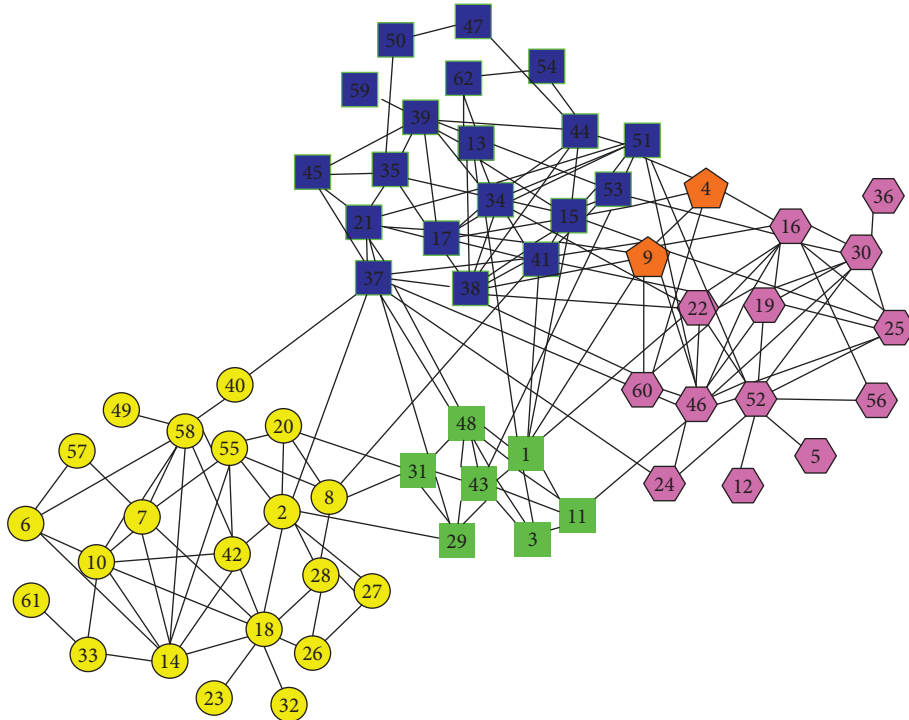FIGURE 5: American College Football network divided into 13 communities.

FIGURE 6: Dolphins network divided into four communities.

while the original network has two communities. But it can be found that the four communities are divided into two subcommunities in two communities in the original network. Node 4 and node 9 are overlapping nodes.

These three experiments are the results of community division, showing a better effect compared with the original network, and there is a higher accuracy rate.

This paper analyzes the modularity Q_c. Shen put forward the criterion for evaluating the closeness of overlapping community, indicating that the greater the Q_c was, the closer the community was; the sparser the community is connected, the better the community division results are. Based on this view, this paper calculates the modularity of the above two algorithms and compares them with other overlapping community detection algorithms in detail.

These algorithms are proposed by Shen et al. [18] and Zhang et al. [19]. Evans and Lambiotte [20] and Duanbing et al. [21] proposed overlapping community detection algorithm, and the results of the experimental comparison are shown in Table 1.

Algorithm one is the overlapping community detection algorithm based on the center node; algorithm two represents the overlapping community detection algorithm based on k-faction.

From Table 1, the overlapping community detection algorithm based on k-faction has a better effect. For the 34-node Club network, the algorithm in the module of Shen, Zhang, and Evans and other algorithms proposed were slightly smaller, and for the Football network and

TABLE 1: Comparison of algorithm modules.

Dataset	34-node Club network	115-node Football network	62-node Dolphins network
Algorithm 1	0.45	0.57	0.51
Algorithm 2	0.42	0.60	0.55
Actual network partitioning	0.42	0.56	0.40
Shen	0.67	—	0.54
Zhang	0.42	0.58	—
Evans	0.46	—	—
Chen	0.38	—	0.33

Dolphins network, this algorithm is better than other algorithms.

Therefore, the overlapping community detection algorithm based on the k-faction has a better effect whether it is from the accuracy of the community division or the angle of the module.

3. Collaborative Filtering Recommendation Algorithm Based on the Community Detection

3.1. Basic Idea. Aiming at the problems in the traditional collaborative filtering recommendation system, a collaborative recommendation algorithm based on community detection is proposed in this paper. In collaborative filtering recommendation algorithms based on the community detection, the users themselves contain their own feature information. User attribute content largely reflects the similar

relationship between friends. At the same time, the user's feature attributes are stable and can reflect the relationship between users. Therefore, in the collaborative filtering recommendation, the user's tag information feature is particularly important.

By introducing the user feature attribute, the similarity of the feature attributes of the user's neighbors can be extracted, and the weight of the unrelated items of the target item in the similar items can be reduced, so that the calculation of the neighbor user set is more accurate. Recommendation items in collaborative filtering recommendation based on community detection are classified, and it is shown in the following equation:

$$
\begin{aligned}
I &= C_1 \cup C_2 \cup \ldots \cup C_K, \\
C_1 &= \left\{ i_{11}, i_{12}, \ldots, i_{1j_i} \right\}, \\
C_K &= \left\{ i_{k1}, i_{k2}, \ldots, i_{kj_k} \right\},
\end{aligned}
\tag{4}
$$

where U represents the set of all users in the system and C_K indicates the kth class.

3.2. User Feature Matrix Construction.

In the case of similar calculations, the scoring value of the improvement factor should be fully considered. Therefore, this paper designs an improved user similarity calculation method based on the collaborative filtering recommendation system. Using the linear combination method, the corresponding weight X is set, and the traditional item H similarity result and the user label category similarity result are combined to jointly serve as the similarity between the two items. The equation is as follows:

$$
\mathrm{sim}\,(i, j) = (1 - \lambda) \times \mathrm{sim}_{R}\,(i, j) + \lambda \times \mathrm{sim}_{\text{item-cate}}\,(i, j), \tag{5}
$$

where $\mathrm{sim}_R\,(i, j)$ is the similarity calculation of the user scoring matrix according to the traditional three similarity measurement methods.

Sarwar used the smallest dataset of MovieLens to compare the three similarities and used MAE as a measurement indicator. The experimental results show that the optimal MAE can be obtained by using the corrected cosine similarity for scoring prediction.

The experiment is similar to the dataset in the paper. Therefore, this paper also uses the modified cosine similarity as the item similar calculation method.

$\mathrm{sim}_{\text{item-cate}}\,(i, j)$ is used to calculate the similarity in the user-category matrix. Since the matrix table is a binary value, it is reasonable to use the cosine similarity calculation. The calculation equation is as follows:

$$
\mathrm{sim}\,(i, j)_{\text{cate}} = \cos\left(\overrightarrow{i}, \overrightarrow{j}\right) = \frac{\overrightarrow{i} \cdot \overrightarrow{j}}{\left\| \overrightarrow{i} \right\| \cdot \left\| \overrightarrow{j} \right\|}, \tag{6}
$$

where $\overrightarrow{i}$ and $\overrightarrow{j}$ are the corresponding row vectors of user i and user j in the user-category matrix, respectively, and indicate the type contained by user i and user j.

Community detection algorithm is used to divide a given target social network into a number of communities that match the actual situation. Each community represents a social circle that is not used by everyone. Potential

collaborative filtering is recommended for target users in these relatively closely related and smaller communities. In this way, the subsequently constructed user-label matrix shrinks the matrix size. Moreover, most of the nodes in the community have similar tag attributes, so the sparseness of the collaborative filtering matrix can be reduced. Construct a user-label scoring matrix using the proposed user-label association strong concept in the community where the target user is located, and calculate the similarity between the user and other users in the community according to the improved similarity calculation formula. The target user's community, the user-label association strength concept, is constructed by using the proposed user-label association strength concept, and the similarity between the user and other users in the community is calculated according to the improved similarity calculation formula. Then, recommend the top K users with the highest similarity as potential friends for the target users.

4. Experiment Design and Discussion

4.1. Experimental Dataset.

In order to test the algorithm of this paper, we have downloaded the classic MovieLens from the Internet and dataset and stored in local database.

MovieLens (http://movielens.umn.edu/) was found in 1997 to a recommendation system, web film based on the number of users, tens of thousands of daily access to the system, and the system can accept user score of the film and movie recommendation list for the user. Currently, the site's users have more than 43000 people. The user rating of the film has more than 3500.

We selected 943 users from the user-rating database and 1682 movies, a total of 100000 scoring information, each of which has at least 20 score information, score from 1 to 5 of the integer; the greater the score, the higher the degree of the user's preference for the film; conversely, the lower.

4.2. Experimental Evaluation Index.

Different similarity indicators will result in different prediction scores. Mean absolute error (MAE) and root mean square error (RMSE) are two common indicators for measuring the accuracy of similarity methods. The smaller the value, the better the prediction accuracy. The usual approach is to divide the original dataset into the training set and test set; then, use the training set to get the prediction model and the test set to evaluate the prediction results, that is, the prediction results and the actual score of the MAE and RMSE values. It is defined as follows:

(1) Mean absolute error (MAE) is the average of the absolute error of the user's predicted score $r'_{u,i}$ and the true score $r_{u,i}$ in the scoring test set:

$$
\mathrm{MAE} = \frac{1}{|p|} \sum_{(u,i) \in p} \left| r_{u,i} - r'_{u,i} \right|. \tag{7}
$$

(2) Root-mean-square error (RMSE) is the mean square root of the true score value r_{ui} and the predicted score value r'_{ui} of the user in the test set:

$$\text{RMSE} = \sqrt{\frac{1}{|p|} \sum_{(u,i) \in p} (r_{ui} - r'_{ui})^2}. \tag{8}$$

$|p|$ in equations (7) and (8) represents the size of the test set.

4.3. Experimental Program. To facilitate the evaluation of performance indicators, the original MovieLens datasets are divided into training sets and test sets. The author randomly selects a part from the user scoring set as the training set, and the remaining scoring data shall become the test set. The training set is used for model training, whereas the test set is used for testing the pros and cons. This paper tests the collaborative filtering algorithm based on the community detection as well as MAE and RMSE indicators of the collaborative filtering algorithm by using Pearson's correlation and cosine similarity. There are two factors significantly affected by the collaborative filtering recommendation: the scarcity of the dataset and the size of neighbor user sets (i.e., the value of k). These two factors are also considered in the validation of the collaborative filtering algorithm.

Based on the above factors, this paper designed the following experimental scheme:

Experiment 1: select different proportions of training sets and test sets in the MovieLens dataset. If the neighbor user set size k is constant, consider comparing the performance of the above algorithms in the case of different data sparsity. Thereby, it is possible to verify the extent to which the recommendation effect is affected under the condition that the system obtains a valid amount of information.

Experiment 2: select a certain percentage of training and test sets in the MovieLens dataset. When the neighbor user set size k is changed, the performance of the above algorithm is compared to verify the influence of the nearest neighbor user set size on the recommendation effect.

In order to avoid the training set and test set of a certain partition, the experimental results are brought by chance. We use the same method to divide the MovieLens dataset into 5 and get 5 sets of different training sets and the corresponding test sets. So, each training set and test set has 5 experiments, and take the average results of the 5 experiments is at the final result of the experiment.

4.4. Experimental Results and Analysis. In order to successfully carry out the experiment, the need is to set some parameters. In this paper, we recommend that the size $|U|$ of the candidate user set U in the algorithm be set to about 30% of the number of users in the training set. For convenience, the collaborative filtering algorithm based on community detection in this paper uses CFCD (collaborative filtering community detection). The cooperative filtering algorithm

based on cosine similarity is expressed by CFC (collaborative filtering cosine). The collaborative filtering algorithm based on Pearson's correlation is expressed by CFP (collaborative filtering Pearson).

Experiment 1: with the MovieLens dataset, the size of the training set is changed if the neighbor user set size k value remains unchanged. Table 2 is based on the premise that the neighbor user set size k is 30. The results of experiments on the proportions of each training set are shown in Table 2.

We draw the line and bar chart corresponding to Table 2. The details are shown in Figures 7 and 8. It can be seen from Figure 7 that the MAE value of the CFCD algorithm proposed in this paper is smaller than that of the CFP algorithm and CFC algorithm. Figure 8 shows that the RMSE value of the CFCD algorithm proposed in this paper is also smaller than that of the CFP algorithm and the CFC algorithm. Thus, the CFCD algorithm is better than the CFP algorithm and CFC algorithm.

Experiment 2: with the MovieLens dataset, in the case of a certain size of the training set, to change the size of the K value of the nearest neighbor user set. Table 3 shows the comparison of MAE and RMSE values at different k values for the training set ratio of 80%.

We draw out Table 3 corresponding to the line chart and bar chart in detail. It is shown in Figures 9 and 10.

Experimental results show the MAE and RMSE values of the collaborative recommendation algorithm based on community detection are smaller than the other two algorithms. The reaction is on the line graph, and in the case of a certain proportion of the training set, the value of k increases. The MAE and RSME polylines of CFCD are below those of CFC and CFP, indicating that the proposed algorithm CFCD is better than the other two algorithms.

When k takes about 30, in fact, the MAE and RSME of the three algorithms are the smallest; when the size of the neighbor user set is 30, the recommended effect is best. Figures 9 and 10 are the corresponding line graphs of Table 3.

It can be seen that CFCD algorithm's MAE and RMSE values are lower than those of the other two algorithms. It can be shown that the CFCD algorithm is better than the CFC and CFP algorithms.

5. Conclusions

This paper introduces social network-related technologies into collaborative recommendation technology and proposes a collaborative recommendation algorithm based on community detection. In order to solve the problems existing in the traditional and recommended technologies, this paper proposes a collaborative recommendation method based on community detection based on community discovery technology and collaborative recommendation technology in social networks. The algorithm

TABLE 2: Experimental results for different training set ratios ($k = 30$).

Ratio of training set (%)	MAE			RMSE		
	CFCD	CFC	CFP	CFCD	CFC	CFP
20	0.814	0.847	0.847	0.995	1.011	1.001
30	0.803	0.848	0.825	0.965	0.995	0.987
40	0.794	0.834	0.814	0.968	0.978	0.968
50	0.780	0.835	0.800	0.917	0.957	0.935
60	0.771	0.830	0.798	0.867	0.934	0.914
70	0.770	0.828	0.787	0.827	0.899	0.867
80	0.765	0.823	0.777	0.812	0.844	0.822
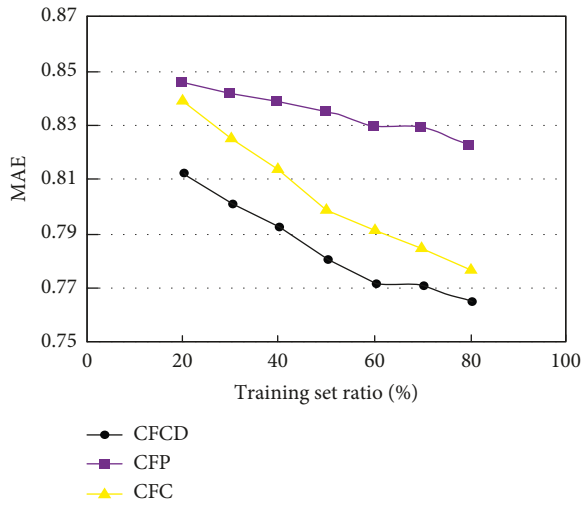
FIGURE 7: MAE results for training set ratio changes ($k = 30$).
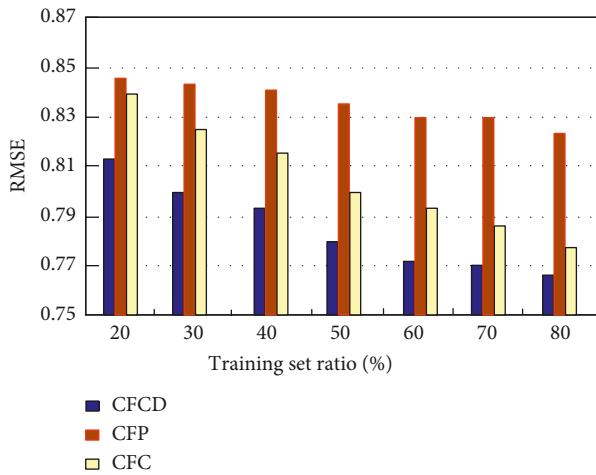
FIGURE 8: RMSE results for training set ratio changes ($k = 30$).

FIGURE 9: K value variation of the MAE results (training set ratio of 80%).
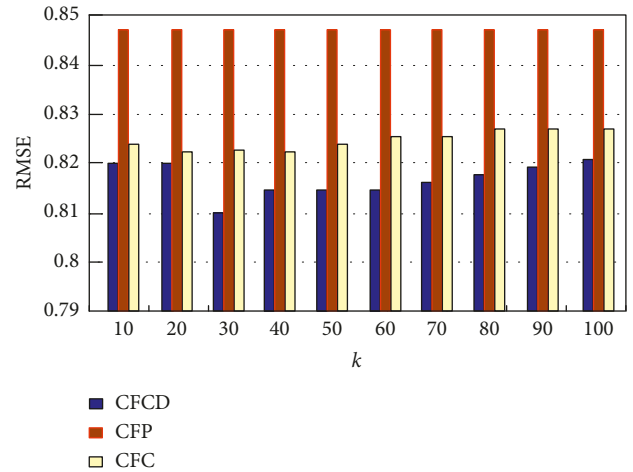
FIGURE 10: K value variation of the RMSE results (training set ratio of 80%).

TABLE 3: Experimental results for the training set ratio of 80%.

K	MAE			RMSE		
	CFCD	CFC	CFP	CFCD	CFC	CFP
10	0.772	0.829	0.708	0.819	0.845	0.823
20	0.767	0.823	0.777	0.812	0.846	0.824
30	0.765	0.824	0.776	0.813	0.845	0.827
40	0.763	0.826	0.778	0.815	0.847	0.823
50	0.771	0.823	0.779	0.816	0.847	0.824
60	0.772	0.826	0.780	0.817	0.848	0.825
70	0.775	0.824	0.781	0.816	0.845	0.826
80	0.778	0.823	0.782	0.817	0.849	0.827
90	0.776	0.824	0.783	0.818	0.846	0.824
100	0.781	0.831	0.783	0.819	0.851	0.827

adopts the method of rough selection of neighboring user sets, which effectively reduce the system calculation time, and at the same time, the scoring preprocessing method is adopted to prevent the problem caused by data sparseness. The performance of the proposed algorithm is verified by experiments. This paper designed two experiments using the MovieLens datas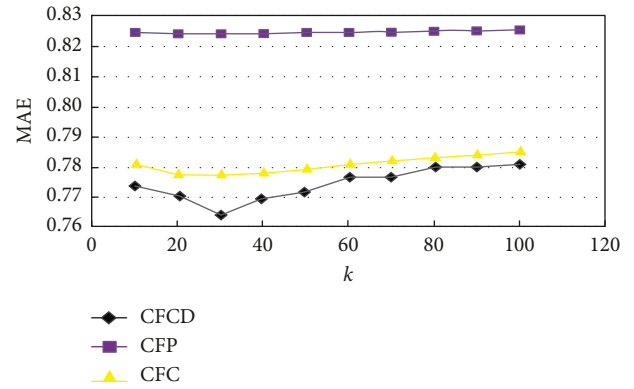et. The performance indicators MAE and RMSE of the community-based collaborative recommendation technology and the collaborative recommendation technique based on Pearson's similarity and cosine similarity are tested separately. The experimental results show that the proposed technique is superior to the performance of collaborative recommendation based on Pearson's similarity and Cosine similarity.

Further research will be carried out in the following areas:

(1) Information acquisition: in addition to obtaining basic user information, it is necessary to dig out more information implied by users. In case, if the new user rating information is insufficient, the user's non-evaluation information may be considered as the user's supplementary information so that the new user can be more accurately recommended.

(2) Recommended technical aspects: in order to improve the accuracy and scalability of the recommendation system and ensure that the system performs real-time recommendation under large-scale data, it is necessary to conduct in-depth exploration and research on the recommended technology.
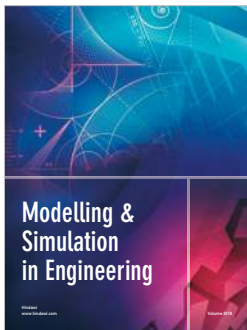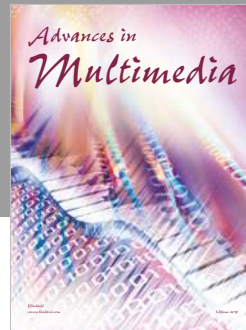
Data Availability

The data used to support the findings of this study are included within the article. Readers can access the data supporting the conclusions of the study from MovieLens (http://movielens.umn.edu/).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Zarzour, F. M. Mohamed, S. Chaouki, and C. Chemam, "An improved collaborative filtering recommendation algorithm for big data," *Computational Intelligence and Its Applications*, pp. 660–668, 2018.

[2] H. Zarzour, Z. A. Sharif, M. A. Ayyoub, and Y. Jararweh, "A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques," in *Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 102–106, Lille, France, April 2018.

[3] Z. Chen, Z. Ning, Q. Xiong, and M. S. Obaidat, "A collaborative filtering recommendation-based scheme for WLANs with differentiated access service," *IEEE Systems Journal*, vol. 12, no. 1, pp. 1004–1014, 2018.

[4] C. Lu, "Recommendation algorithm based on collaborative filtering under social network environment," *BioTechnology: An Indian Journal*, vol. 10, no. 13, pp. 7087–7092, 2014.

[5] Q.-J. Tan, H. Wu, C. Wang, and Q. Guo, "A personalized hybrid recommendation strategy based on user behaviors and its application," in *Proceedings of the 2017 International Conference on Security, Pattern Analysis, and Cybernetics*, pp. 181–186, SPAC, Shenzhen, China, December 2017.

[6] L. Guo, J. LiangYing, Z.Y. Luo, and L. Sun, "Collaborative filtering recommendation based on trust and emotion," *Journal of Intelligent Information Systems*, pp. 1–23, 2018.

[7] J. Aranda, I. Givoni, J. Handcock, and D. Tarlow, *An Online Social Network-Based Recommendation System*, University of Toronto, Ontario, Canada, 2007.

[8] J. Stan, F. Muhlenbach, and C. Largeron, "Recommender systems using social network analysis: challenges and future trends," in *Encyclopedia of Social Network Analysis and Mining*, pp. 1–22, Springer, New York, NY, USA, 2014.

[9] K. G. Saranya and G. Sudha Sadasivam, "Personalized news article recommendation with novelty using collaborative filtering based rough set theory," *Mobile Networks and Applications*, vol. 22, no. 4, pp. 719–729, 2017.

[10] L. O. Colombo-Mendoza, R. Valencia-García, R. Colomo-Palacios, and G. Alor-Hernández, "A knowledge-based multi-criteria collaborative filtering approach for discovering services in mobile cloud computing platforms," *Journal of Intelligent Information Systems*, pp. 1–25, 2018.

[11] H.-G. Ko, I.-Y. Ko, and D. Lee, "Multi-criteria matrix localization and integration for personalized collaborative filtering in IoT environments," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4697–4730, 2018.

[12] A. M. Rashid, K. L. Shyong, and K.L. George, "A highly scalable hybrid model & memory-based CF algorithm," in *Proceeding of the WEBKDD of ACM*, Macau, China, June 2014.

[13] G.-R. Xue, C. Lin, Q. Yang et al., "Scalable collaborative filtering using cluster-based smoothing," in *Proceedings of the 20th Annual International ACM SIGIR Conference*, pp. 15–19, Tampere, Finland, August 2015.

[14] A. Clauset, "Finding local community structure in networks," *Physical Review E*, vol. 72, no. 2, pp. 56–70, 2015.

[15] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[16] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2012.

[17] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, pp. 25–32, 2014.

[18] H. W. Shen, X.-Q. Cheng, and J. F. Guo, "Quantifying and identifying the overlapping community structure in networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 7, Article ID P07042, 42 pages, 2009.

[19] S. Zhang, R.-S. Wang, and X.-S. zhang, "Identification of overlapping community structure in complex networks using fuzzy-means clustering," *Physica A: Statistical Mechanics and Its Applications*, vol. 374, no. 1, pp. 483–490, 2007.

[20] T. S. Evans and R. Lambiotte, "Line graphs, link partitions, and overlapping communities," *Physical Review*, vol. 80, no. 1, pp. 145–148, 2009.

[21] C. Duanbing, Y. Fu, and M. S. Shang, "An efficient algorithm for overlapping community detection in complex," *Global Congress on Intelligent Systems*, pp. 244–247, 2009.