User Preference-oriented Collaborative Recommendation Algorithm in E-commerce

Huiying Gao
Beijing Institute of Technology/School of management and economic, Beijing, China
Email: huiying@bit.edu.cn

Susu Wang, Bofei Yang, and Hangzhou Yang
Beijing Institute of Technology/School of management and economic, Beijing, China

***Abstract*—Collaborative recommendation is a key issue today in e-commerce, which helps users find the information of products which they are interested in from the mass of information. Aiming at the problem in an insufficient personalization of e-commerce recommendation system， a user preference-oriented network model is established in the paper. Related theories and techniques of clustering are used to analyze the extracted preferences of the users. Finally, combined with the traditional collaborative filtering method, the paper proposed the user preference-oriented collaborative recommendation algorithm. The experimental results show that compared to the traditional collaborative filtering algorithm, user preference-oriented collaborative recommendation algorithm can effectively improve the efficiency of e-commerce recommendation system in ensuring the accuracy of the premise.**

***Index Terms*—collaborative recommendation, user preference, e-commerce, clustering analysis, neural network**

I. INTRODUCTION

With the popularity of computer and Internet, as well as the continuous improvement of supply chain and modern logistics, e-commerce has got unprecedented development and has gradually become an indispensable part of people's life. At the same time, some problems existing in current e-commerce arise, such as the conflict between the rapid expansion of product information and the personalized user demands. On the one hand, along with the increase of the online shops the merchants are providing variety of goods. However, users may be at a loss in the massive commodity information, so they could not find the goods and information which they really need. On the other hand, the users' demands is becoming diversified and personalized, so that the user's satisfaction may decrease, if businesses could not target and mine the potential demand of users accurately. Therefore, how well the e-commerce websites accurately segment the customers, meet the personalized needs of customers and provide right services for them so as to enhance the customer loyalty and its profitability, which provide personalized services for the users. Personalization[1] can be expressed as: personalization is a process of automatically adaption, restructuring and representation of the related information for each individual member. In a nutshell, personalization is the ability to provide users with targeted contents and services based on the knowledge of users' preferences and behaviors.

In the paper, a user preference-oriented network model was established based on users' interests. The extracted user preferences were clustered and analyzed by using the related theories and techniques of network analysis and clustering. Finally, combined with the traditional collaborative filtering method, a user preference-oriented e-commerce collaborative recommendation algorithm was proposed. At the end of the paper, the experimental results was shown and analyzed which indicate that user preference-oriented e-commerce collaborative recommendation algorithm can improve the efficiency of e-commerce recommendation system more effectively than traditional collaborative filtering algorithm in ensuring the accuracy of the premise.

Corresponding author: Huiying Gao, Associated Professor
Email: huiying@bit.edu.cn

II. STATE OF ART

A. User Preference

User preference [2] refers to the users' rational and slant choices when they consider the goods and services. It is an integrated result of users' cognition, psychological feelings and rational economics balance.

Nowadays Internet community in which the dominant or recessive network forms is increasingly arousing attention. Scholars have done some research on using the network to get the user preference to form a network in which users' preferences is similar. Nisgav[3] etc. considered how a user find its similar ones in the social network and put forward to judge whether the two users are similar according to the answer of the question. Yuan Guan[4] etc. combined two kinds of recommendation algorithm and considered the user parameters into the recommendation system, emphasizing the importance of the heterogeneity among users. They estimated the current users' preferences according to the similar uses' preferences. If a user is closely linked to other certain

doi:10.4304/jsw.9.7.1886-1893

users, it can be infered that its preference is similar to that of the users. Saha and Getoor[5] proposed community recommendation algorithm similarity that other communities can be recommended to members of the community by calculating the similarity among groups based on group. Mishra etc.[6] conducted community recommendation through group clustering the social network. Such recommendations treated the groups in the community as independent study objects and applied the methods of corresponding similarity calculation and clustering etc. to recommend the research object. Some scholars have found the big data hiding behind the social platforms such as blogs and weibo in the era of Web 2.0, and have done some research based on them. For example, Chen Jun[7] proposed a collaborative filtering model based on the Web social network for the blog recommendation. Chen Jun etc.[8] put forward a new model of personalized web information recommendation based on web social network by analyzing users' preferences, the generation of web social network and the collection of web information that should be filtered. In the approach of personalization for group learning supporting proposed by Xin Wan etc.[9], learners' relationship was extracted based on learning processes and learning activities using natural language process technologies and data mining.

B. Collaborative Recommendation

Collaborative filtering recommendation technology is one of the most successful recommendation technologies so far. Its basic idea is to get the preference information of target user and compare it to that of other users to look for the similar group of the target user. So that the target user's prediction rating of the resource can be predicted by the similar groups' prediction rating of the resource and recommendation can be made for target user based on it.

The basic process of collaborative recommendation includes three main steps: (1)the representation of user information; (2)selection of neighbor users; (3) generation of recommendation results.

User information is mainly the user's rating data which can be divided into explicit rating and implicit rating. Explicit rating is the numerical rating on some items that users input into the recommendation system, while the implicit rating is the prediction rating on items predicted by analyzing the character of users' behaviors according to users' history of browsing and consumption. However, whether explicit rating or implicit rating, user rating information is usually represented by rating vector in which each element is on behalf of one user's rating on corresponding item. All the user information make up the user-item rating matrix that can be seen in table I in which the number at row i, column j represents user i 's rating on item j .

TABLE I.
USER-ITEM RATING MATRIX

USER	ITEM				
	I_1	I_2	…	I_{n-1}	I_n
U_1	4	5	…	0	1
U_2	3	5	…	5	1
…	…	…	…	…	…
U_{m-1}	2	5	…	4	5
U_m	4	1	…	5	4

The selection of the nearest neighbor which determines the quality of the collaborative filtering algorithm is the core of collaborative filtering. There are mainly three calculation method of similarity: cosine similarity , pearson and the modified cosine similarity[10]. Their formulas to calculate are as follow:

Cosine similarity:

$$sim(\vec{i},\vec{j})=\cos(\vec{i},\vec{j})=\frac{\vec{i}\cdot\vec{j}}{\|\vec{i}\|\cdot\|\vec{j}\|} \tag{1}$$

While user i 's and user j 's n-dimensional rating vector are $\vec{i}$ and $\vec{j}$, and $sim(\vec{i},\vec{j})$ represents the similarity between user i and user j .

Pearson similarity:

$$sim(\vec{i},\vec{j})=\frac{\sum_{c\in I_{ij}}(r_{i,c}-\bar{r}_i)\cdot(r_{j,c}-\bar{r}_j)}{\sqrt{\sum_{c\in I_{ij}}(r_{i,c}-\bar{r}_i)^2}\cdot\sqrt{\sum_{c\in I_{ij}}(r_{j,c}-\bar{r}_j)^2}} \tag{2}$$

in which user i 's and user j 's n-dimensional rating vectors are $\vec{i}$ and $\vec{j}$, rating vectors of all users constitute the user-item rating matrix $R_{m\times n}$, the item set of i and j 's common rating is represented by $I_{i,j}$, $r_{i,c}$ and $r_{j,c}$ are respectively user i 's and user j 's rating on item c , $\bar{r}_i$ and $\bar{r}_j$ are respectively user i 's and user j 's mean rating of item ratings that they have scored.

Modified cosine similarity:

$$P_{p,i}=\bar{r}_p+\frac{\sum_{q\in N_p}sim(p,q)\cdot(r_{q,i}-\bar{r}_q)}{\sum_{q\in N_p}sim(p,q)} \tag{3}$$

in which $\bar{r}_p$ is the mean rating of all items that the target user p has scored, $r_{q,i}$ is the real rating that neighbor q scored on item i , while $\bar{r}_q$ is the mean rating of items that have been scored by the neighbor q .

Recommendation result is generated based on the prediction rating on items that have not been scored by target user.

In the field of e-commerce recommendation, Kamak [11] proposed a fuzzy computing model of trust and credibility system, which has promoted the effect of recommendation system by the method of double filtration. Altingovde[12] proposed a clustering search strategy for collaborative recommender system. Sarah K.Tyler [13] put forward that the data of social network and historical rating should be combined to achieve product recommendation. Lu J[14] presented a hybrid fuzzy semantic recommendation (HFSR) approach which combined techniques of item-based fuzzy semantic similarity and item-based fuzzy collaborative filtering (CF) similarity. Relevant business partners can be recommended to personal business user through this method. Lü L [15] compared and evaluated the existing recommendation system algorithms, examining their role in the future development. The potential impact and future direction of development were also discussed. In order to solve the problems of scalability and sparsity in the collaborative filtering, Li[16] proposed a personalization recommendation algorithm based on rough set, and it uses a new similarity measure to find the target users' neighbors. Gong[17] presented a recommendation algorithm on integration of item semantic similarity and item rating similarity, which help alleviate the sparsity problem in e-commerce recommender systems.

Scholars have carried out some research on applying user preference network in e-commerce recommendation. It depends on not only their own preferences but also the influence of friends when users make purchase decisions. Chua[18] focused on the social relevance modeling, and proposed a social relevance framework to predict items that users may select. Using the nearest neighbor algorithm, Gan[19] established user similarity network by user's historical data, and a collaborative filtering algorithm was proposed based on network to overcome the adverse effects of hot items. It reached a reasonable balance between accuracy and diversity. Shinde[20] proposed a fast k-medoids clustering algorithm for recommender system, in which opinions from the users are collected in the form of user-item rating matrix.

The research on the users' preference network and the recommendation method in e-commerce are still inadequate. For the formation of user preference networks in the existed research work were mostly not in e-commerce community, so it needs further consideration on user preference in e-commerce in order to reflect the user's true preference.

Collaborative filtering is suitable for groups that interact with system, it has been widely applied in e-commerce recommendation field and has achieved remarkable results. The traditional collaborative filtering algorithm searches for neighbor users in a huge users group by analyzing user-item scoring matrix. There is a great efficiency bottleneck in the algorithm when the data is excessive.

Therefore, on the basis of previous studies, this paper proposes a user preference oriented e-commerce collaborative recommendation algorithm. Based on historical behavior or user preference similarity, hidden links were found to build a potential interest network of a user. To find the similarity between users, we start with user's interest and preference. Preference network is found based on user's preference link. On this basis, hierarchical clustering algorithm is used to cluster users and finally a user preference oriented e-commerce collaborative recommendation algorithm is formed combined with traditional collaborative filtering algorithm.

III. User Preference Network

User Modeling is the foundation of e-commerce collaborative recommendation. In our method, user preference is extracted by analyzing the user's explicit scoring history. The preference of user i is represented by vector $U_i = (u_{i1}, u_{i2}, \ldots, u_{in})$, in which $u_{ij} \in [0,1]$ means the preference degree of user i to concept j, and n is the total number of concepts. Similarly, the recommended item V_k in recommendation system is described by concept weight vector $(d_{k1}, d_{k2}, \ldots, d_{kn})$, and d_{kj} means the weight of item V_k on concept j. This makes user preference and items in the same vector space, and it is easy to extract user preference from the scoring history of users and to calculate user's degree of interests on each concept.

We believe that the user is interested in the item when he or she scores it, so user preference is extracted from the user's scoring history. According to the above settings, p_{ij} is used to represent the preference degree of user i to concept j, D_i is the set that user i have scored in item V_k. The degree of preference can be obtained from the sum of corresponding concept weight of item set that user has scored. Then the max-min method [21] is adopted to normalize it and user's preference weight u_{ij} is reached.

$$p_{ij} = \sum_{k \in D_i} d_{kj} \tag{4}$$

$$u_{ij} = \frac{p_{ij} - \min\limits_{1 \le j \le n}\{p_{ij}\}}{\max\limits_{1 \le j \le n}\{p_{ij}\} - \min\limits_{1 \le j \le n}\{p_{ij}\}} \tag{5}$$

We can draw by the Equation (5) that: if $p_{ij} = \max\limits_{1 \le j \le n}\{p_{ij}\}$, then $u_{ij} = 1$ and if $p_{ij} = \min\limits_{1 \le j \le n}\{p_{ij}\}$, then $u_{ij} = 0$. It means the degree of user's preference to different items is increasing while u_{ij} is increasing from 0 to 1.

The similarity between two users can be calculated according to the user's preference weight. The relationship of different users in preference network is founded based on the similarity among users. The similarity between user U_x and user U_y is calculated by the cosine method:

$$sim(U_x, U_y) = \cos(U_x, U_y) = \frac{U_x \cdot U_y}{\|U_x\| \cdot \|U_y\|} \quad (6)$$

The value of similarity degree is namely link strength degree of user node in preference network. And we named it the user preference network.

IV. User Preference Oriented E-Commerce Collaborative Recommendation Algorithm

This paper proposed user preference oriented e-commerce collaborative recommendation algorithm based on user preference that is extracted above. User is clustered based on the user preference network. We can find the nearest neighbor from the cluster that target user is in. Then recommendation is generated based on traditional collaborative filtering algorithms.

A. The Basic Process of Algorithm

The basic process of algorithm is showed in figure 1.

Step 1, user preference is extracted based on user's scoring history and the user preference network is founded based on it in which the attachment of the network represents the similarity of them.

Step 2, user nodes in the network are clustered and the cluster that target user is in is selected as the candidate neighbor users when target user needs to be recommended something.

Step 3, K-Nearest neighbor algorithm (KNN) method is adopted to select the top n^k nearest users with target user as the neighbor users, and that is based on the similarity between target user and all the other candidate users that is calculated according to the user-item matrix.

Step 4, recommend for target user based on the top n^k neighbor users. According to the top n^k neighbor users' scores, items on which target user has no scoring history can be given a score based on them.

As the recommendation of traditional collaboration recommendation algorithm is based on all users in the database whom have scored, it makes the time that recommendation takes much longer.

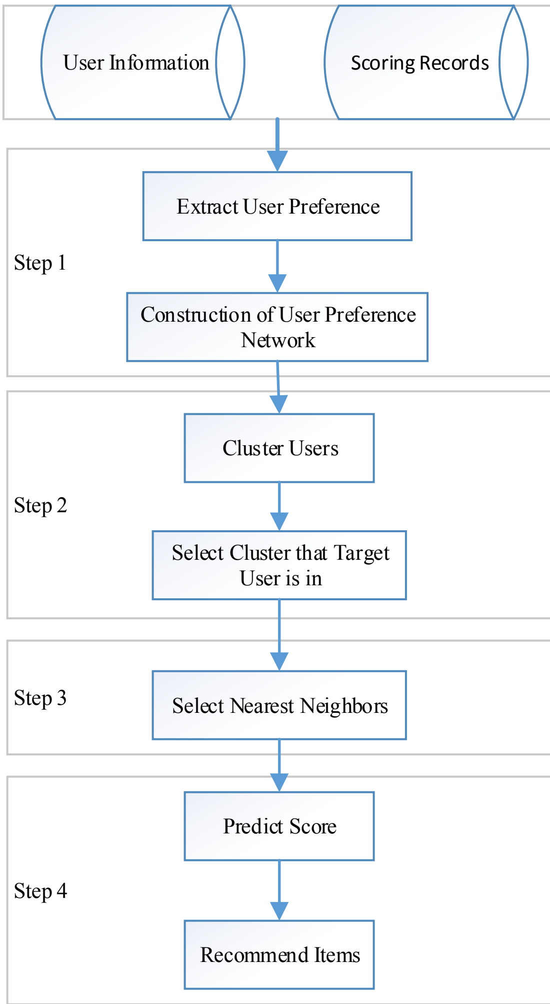

Fig. 1 The processes of user preference-oriented collaborative recommendation algorithm

User preference-oriented collaborative recommendation algorithm combined both memory based and model based method. That is, combined the nearest neighbor recommendation method and user clustering together. Of which the establishment of preference network and user clustering can be taken offline. When new users, items or scorings come, the system can recalculate recommendation in idle time to avoid consuming more resources, and it is beneficial for system to make a real time recommendation. By selecting the nearest neighbor of target user, it reduces the range of neighbor users for real time recommendation and consequently saved computation time. On the other hand, it precisely locates similar users in preference and help to improve the quality of recommendation.

B. Realization of Algorithm

Based on the user preference network established in section 2, bottom-up hierarchical clustering algorithm is used to cluster users. The hierarchical clustering algorithm organizes data into several groups and forms a corresponding tree diagram to cluster. This method not only can suit any type and any attribute data set, but also can flexibly control clustering size of different levels.

And the number of classifications needs not to be determined in advance, so it has a strong ability to cluster. The basic idea of bottom-up hierarchical clustering algorithm is: First, the n nodes in the network are respectively seen as n independent clusters, and the link similarity between any two clusters is calculated. The similarity is ordered in descending order and the two largest clusters are merged. In this way, all clusters are merged into a big cluster in the end and thus a tree diagram is formed. Then according to specific criteria, a certain height is selected to cut to obtain the final clustering result. The algorithm process can be represented by the tree diagram shown in Figure 2. Cutting at different levels will get different results.

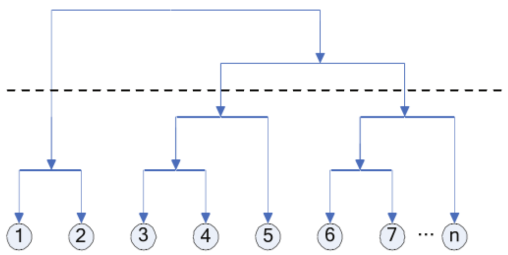

Fig. 2 The tree view of algorithm of hierarchical clustering

Based on the result of user clustering, m clusters of users $\{C_1, C_2, \ldots, C_m\}$ are obtained. Assuming the target user $i \in C_t$, a subset of the user-item rating matrix represented by a matrix R_t ranking $l \times s$ is got, where row l represents there are l users in the cluster C_t and column s represents there are s items, the element at row i and column j represents that the scoring of user i on item j, it is 0 when there is no scoring.

$$R_t = \begin{bmatrix} r_{t_1 1} & r_{t_1 2} & \cdots & r_{t_1 s} \\ r_{t_2 1} & r_{t_2 2} & \cdots & r_{t_2 s} \\ \cdots & \cdots & \cdots & \cdots \\ r_{t_l 1} & r_{t_l 2} & \cdots & r_{t_l s} \end{bmatrix} \tag{7}$$

Based on the user-item scoring matrix, cosine similarity showed in Equation (6) is adopted to measure the similarity between users. If every row L_i of the matrix is the scoring vector of users, then the similarity of user U_{t_x} and user U_{t_x} is calculated as shown in Equation (8):

$$sim\left(U_{t_x}, U_{t_y}\right) = \cos\left(L_{t_x}, L_{t_y}\right) = \frac{L_{t_x} \cdot L_{t_y}}{\left\|L_{t_x}\right\| \cdot \left\|L_{t_y}\right\|} \tag{8}$$

After the similarities between different users are calculated, the top n^k users that have highest similarities with target user i is selected as its neighbors, and is represented by N_i. Prediction score on item that target user not scored before is done to generate recommendation. The prediction score pre_{ij} that user i made on item j is calculated according to Equation (9) [22]:

$$pre_{ij} = \overline{r_i} + \frac{\sum_{u \in N_i} sim(i,u) \cdot \left(r_{uj} - \overline{r_u}\right)}{\sum_{u \in N_i} \left|sim(i,u)\right|} \tag{9}$$

In which $sim(i,u)$ represents the similarity between target user and its neighbor, and r_{uj} represents the scoring of user u to item j. $\overline{r_i}$, $\overline{r_u}$ respectively represents the average score that user i and user u to the item. A number of highest scores are selected as recommendation results back to the current user after the prediction.

V. Algorithm Verification

A. Data Source

In order to verify the user preference-oriented e-commerce collaborative recommendation algorithms and processes proposed in this paper, we conducted the verification based on the Movielens data set.

Movielens(http://movielens.umn.edu) is a recommendation system based on Web for research, which recommends movies through users' rating (total score is 5). Two rating datasets (http://www.grouplens.org/data/) are developed and published by the university of Minnesota in the United States, one of which includes 943 users' 100000 ratings on 1682 movies, each user scores at least 20 movies and every movie belongs to one of the 19 kinds of movies. The other one contains 6040 users' 1000209 ratings on 3952 movies.

According to the actual situation, we randomly filtered out 943 users' 99965 videos rating records from Movielens data set to do our research.

B. Running Environment

The experiment is running on a personal computer, and its configurations are as follow:

1) *Hardware Configuration*

CPU: intel Core i3-2330M
CPU Speed: 2.2GHz
Three cache: 3M
Memory: 2GB
HDD: 500GB

2) *Software Configuration*

Operation System: Microsoft Windows 7
Software being used: Eclipse 3.7.1, mySQL 5.5.25 and SPSS 20.0

C. Result and Analysis

The proposed algorithm was experimented on the data and compared with traditional collaborative recommendation algorithm using the evaluation indicator MAE. Figure 3 shows example of the expression of user preference.

1	0.0000	0.5088	0.2807	0.1228	0.2456	0.8947	0.1930	0.0877	1.0000	0.0175	0.0175	0.0702
2	0.0000	0.1739	0.0435	0.0435	0.0870	0.3913	0.2609	0.0000	1.0000	0.0000	0.0435	0.0435
3	0.0000	0.5000	0.2500	0.0000	0.0000	0.5000	0.3333	0.0833	1.0000	0.0000	0.0833	0.1667
4	0.0000	1.0000	0.5000	0.0000	0.0000	0.3333	0.5000	0.1667	0.5000	0.0000	0.0000	0.0000
5	0.0000	0.7750	0.5000	0.2250	0.3250	1.0000	0.1000	0.0000	0.3500	0.0000	0.0000	0.3750
6	0.0000	0.3404	0.2979	0.1489	0.2766	0.8298	0.1915	0.0000	1.0000	0.0426	0.0638	0.0426
7	0.0000	0.5244	0.2805	0.0610	0.1951	0.5488	0.2195	0.0244	1.0000	0.0244	0.0366	0.3415
8	0.0000	1.0000	0.5000	0.0000	0.0909	0.2273	0.2727	0.0000	0.2273	0.0000	0.0000	0.0000
9	0.0000	0.5000	0.5000	0.0000	0.0000	0.8333	0.0000	0.0000	0.8333	0.0000	0.0000	0.1667
10	0.0000	0.2683	0.1463	0.0488	0.0488	0.6098	0.1951	0.0488	1.0000	0.0000	0.0976	0.0244
11	0.0000	0.3600	0.2200	0.0000	0.0600	0.8600	0.1800	0.0000	1.0000	0.0200	0.0200	0.0400
12	0.0000	0.7000	0.2000	0.0000	0.1000	1.0000	0.1000	0.0000	1.0000	0.0000	0.0000	0.0000
13	0.0000	0.6861	0.2701	0.0511	0.1752	0.7372	0.2190	0.0438	1.0000	0.0292	0.0657	0.3285
14	0.0000	0.5333	0.3333	0.2667	0.2000	0.9333	0.2667	0.0667	1.0000	0.0667	0.2000	0.0667
15	0.0000	0.6250	0.2917	0.0000	0.1250	0.7500	0.1250	0.0000	1.0000	0.0833	0.0833	0.0417
16	0.0000	0.5000	0.1923	0.2308	0.2308	0.8846	0.3846	0.0000	1.0000	0.0385	0.1154	0.1538
17	0.0000	0.2727	0.2727	0.0000	0.0909	0.4545	0.1818	0.0000	1.0000	0.0000	0.0000	0.0000
18	0.0000	0.1829	0.1220	0.0732	0.1585	0.7439	0.1098	0.0366	1.0000	0.0366	0.0244	0.0244

Fig. 3 Example of the expression of user preference

In order to facilitate the assessment of performance indicator, we divide the dataset into training data and test set. The training set is used to extract information of user preference and cluster users. The prediction score derived from training set is compared with that of test set. After randomly selecting 80000 ratings from rating set to constitute the training set, rest of it constitute the test set. Test set is used for model training, and training set is for testing the performance of the method.

After testing, by integrating the processing results of methods k-means and hierarchical clustering in software SPSS, we chose the hierarchical clustering which can cluster the test data more evenly. For the determination of the classification number, we clustered 943 users into 19 classes using the bottom-up hierarchical clustering method, according to the advice of Lehmann(1979) that classification number should be between $n/30$ and $n/60$ (in which n represents the number of samples).

Figure 4 shows part of the clustering process. The ordinate represents the number of 943 users, and the abscissa represents 942 clustering steps. If the abscissa of three connected rectangles is equal, it means that the two users represented by the both sides of the columnar bar clustering into one class in this step.

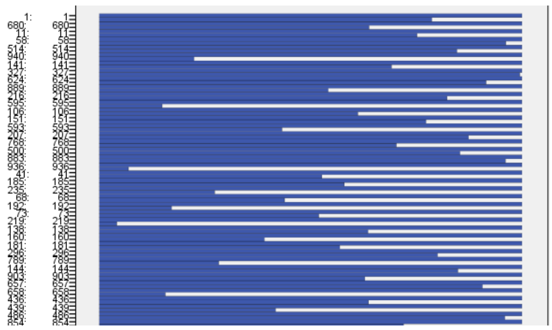

Fig. 4 Part of the process of clustering

Figure 5 shows the distribution of the clustering results, where the abscissa indicates the 19 categories, and the ordinate represents the number of users that each category included. As can be seen from this figure, the minimum class has 13 users, while the maximum has 143 users.

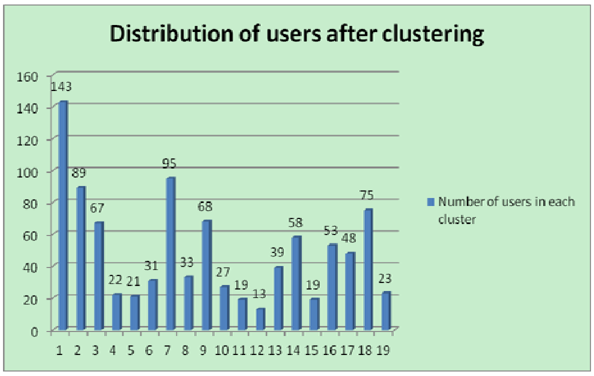

Fig. 5 The result of hierarchical clustering

After clustering, the selection set of the nearest neighbor will change, so that the nearest neighbor calculation results of each user will be different. Figure 6 shows the change of the nearest neighbor of the user with the number "1" after the clustering. Figure 6(a) shows the result of 10 nearest neighbors which is selected from all users of user with the number "1"; Figure 6(b) shows the result of 10 nearest neighbors which is selected from the class that the user belongs to of user with the number "1" after clustering. Comparison shows that four users represented by the circle with the dark color are the same, as the other six users changed after clustering. It can be seen that although finding range is narrower after clustering, the accuracy of looking for the nearest neighbor is ensured to a certain degree.

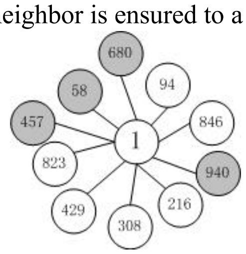

Fig. 6(a) Global nearest neighbor clustering

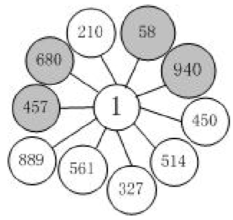

Fig. 6(b) Nearest neighbor after

For all the clusters that include less than 10 users, calculate using the global user-item scoring matrix. Set the number of neighbors (i.e. the value of n^k) 10, and compare the results of the new methods and traditional collaborative filtering algorithm on the dataset.

There are several methods of measuring the accuracy of prediction. Mean Absolute Error (MAE)[23] is a common method which measures the accuracy by calculating the absolute value of a difference between forecasted user ratings and actual user ratings. Let forecasted user rating set be represented by $\{p_1, p_2, ..., p_N\}$, and the corresponding actual rating set be $\{q_1, q_2, ..., q_N\}$, then the Mean Absolute Error MAE is defined as

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \tag{10}$$

Normalized Mean Absolute Error (NMAE)

$$NMAE = \frac{MAE}{q_{\max} - q_{\min}} \quad (11)$$

Root of the Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(p_i - q_i)^2}{N}} \quad (12)$$

The paper chose the Mean Absolute Error (MAE), Normalized Mean Absolute Error (NMAE), Root of the Mean Square Error (RMSE) as the evaluation measure of quality of the recommendation system and used the clustering function of SPSS and calculation function of Matlab to calculate the evaluation index. The results are shown in Table II.

TABLE II.
COMPARISON OF ACCURACY BETWEEN TRADITIONAL AND CLUSTERING ALGORITHMS

	Traditional algorithm	Clustering algorithm
MAE	0.7768	0.7796
NMAE	0.1942	0.1942
RMSE	0.9849	0.9851

Calculate and compare the time of the different part of two algorithms, which is the time spent on looking for the nearest neighbors. In order to eliminate the impact of difference of experiment environment, calculate means of ten results. Then the time the traditional algorithm takes is 0.043 seconds, while that of the clustering algorithm is 0.000037167. Therefore, the proposed user preference-oriented e-commerce collaborative recommendation algorithm greatly reduces the calculating time. In practical application, the number of users and ratings may be massive, so the time clustering algorithm saves will also be considerable, which has certain practical significance to recommendation systems. We can see from Table II that although clustering algorithm may be not as to traditional algorithm on the aspect of recommendation accuracy, the overall difference is small. That is to say compare to traditional algorithm, clustering algorithm ensures the accuracy of recommendation while improves the efficiency of it greatly.

VI. CONCLUSION

With the popularity of e-commerce, online shopping has become an indispensable overhead in our modern life. Whether you buy real stuff or services, the way which e-commerce records satisfaction of users is usually rating and comment. Take the film as an example, China as the country that has the most audiences that see films in cinemas, its accumulation in the film rating data is very valuable data resource in the age of big data. Therefore, how to make use of the rating data to promote consumption is a hot issue for the e-commerce.

This paper puts forward a user preference-oriented collaborative recommendation algorithm in e-commerce. It extracts user preference from users' rating history to establish the preference network model with the user similarity as its edge, and then gets preliminary similar users through clustering, and combines with the traditional collaborative filtering algorithm to recommend to the users. Finally it is verified on the movie rating data. User preference-oriented collaborative recommendation algorithm enhances the real-time performance of the system and improves the scalability of traditional algorithm problem.

In addition, this method also has some certain insufficiency. In the process of extracting user preference, there may be semantic ambiguity. Subsequent research should further provide semantic support for concept space to modify the description of user preference information. This will make user preference model more clear, so as to support the recommendation better. In addition, the method does not consider users' profiles and the sparse problem of rating data. These problems need to be solved in the future study.

ACKNOWLEDGMENTS

This work was supported in part by the Beijing Natural Science Foundation (Grant No. 9133020) and Beijing Institute of Technology (Grant No. 3210012211218).

REFERENCES

[1] Perugini S, Ramakrishnan N. Personalizing interactions with information systems[J]. Advances in Computers, 2003, 57: 323-382.
[2] Mobile user behavior research. http://labs.chinamobile.com/mblog/382108_49663
[3] Nisgav A, Patt-Shamir B. Finding similar users in social networks[C]//Proceedings of the twenty-first annual symposium on Parallelism in algorithms and architectures. ACM, 2009: 169-177.
[4] Guan Y, Zhao D, Zeng A, et al. Preference of online users and personalized recommendations[J]. Physica A: Statistical Mechanics and its Applications, 2013.
[5] Saha B, Getoor L. Group proximity measure for recommending groups in online social networks[J]. networks, 2008, 1(6): 5.
[6] MISHRA N, SCHREIBER R, STANTON I,et al.Clustering the social networks [C]//WAW2007: 5thWorkshop onAlgorithms and Models for theWeb-Graph, LNCS 4863. Berlin: Springer-Verlag, 2007: 56-67.
[7] Chen Jun. Research on collaborative filtering model based on Web social network [D]. Southwest university, 2006.
[8] Chen Jun, Tang Yan. A personalized Web information recommendation model based on Web social network[J]. Journal of computer science, 2006 (4) : 185-193.
[9] Wan X, Jamaliding Q, Okamoto T. Analyzing Learners' Relationship to Improve the Quality of Recommender System for Group Learning Support[J]. Journal of Computers, 2011, 6(2): 254-262.
[10] Qing Hai. The research of core technology of the e-commerce recommendation system[D][D].Beijing university of technology,2009.
[11] Bharadwaj K K, Al-Shamri M Y H. Fuzzy computational models for trust and reputation systems[J]. Electronic Commerce Research and Applications, 2009, 8(1): 37-47.
[12] Altingovde I S, Subakan Ö N, Ulusoy Ö. Cluster searching strategies for collaborative recommendation systems[J].

Information Processing & Management, 2013, 49(3): 688-697.
[13] Tyler S K, Zhang Y. Open domain recommendation: Social networks and collaborative filtering[M]//Advanced Data Mining and Applications. Springer Berlin Heidelberg, 2008: 330-341.
[14] Lu J, Shambour Q, Xu Y, et al. A WEB - BASED PERSONALIZED BUSINESS PARTNER RECOMMENDATION SYSTEM USING FUZZY SEMANTIC TECHNIQUES[J]. Computational Intelligence, 2012, 29(1): 37-69.
[15] Lü L, Medo M, Yeung C H, et al. Recommender systems[J]. Physics Reports, 2012, 519(1): 1-49.
[16] Li H, Zhang S, Wang X. A Personalization Recommendation Algorithm for E-Commerce[J]. Journal of Software, 2013, 8(1): 176-183.
[17] Gong Songjie. A Personalized Recommendation Algorithm on Integration of Item Semantic Similarity and Item Rating Similarity[J]. Journal of Computers, 2011, 6(5): 1047-1054.
[18] Chua F, Lauw H, Lim E. Generative models for item adoptions using social correlation[J]. Knowledge and Data Engineering, 2012, 25(9): 2036 – 2048.
[19] Gan M, Jiang R. Constructing a user similarity network to remove adverse influence of popular objects for personalized recommendation[J]. Expert Systems with Applications, 2013, 40(10): 4044–4053.
[20] Shinde S K, Kulkarni U V. Hybrid Personalized Recommender System Using Fast K-medoids Clustering Algorithm[J]. Journal of Advances in Information Technology, 2011, 2(3): 152-158.
[21] Xu Zeshui. Method and application of uncertain multiple attribute decision making [M]. Beijing: tsinghua university press, 2004.
[22] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[C]//Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1998: 43-52.
[23] Li Cong, Liang Changyong, Ma Li. A collaborative filtering recommendation algorithm based on domain nearest neighbor[J]. Journal of Computer Research and Development, 2008, 45(9): 1532-1538.

Huiying Gao, Dr. Associated professor, was born in Shandong Province, China, in 1976. She received doctor degree in Management Science and Engineering from Beijing Institute of Technology in 2003.

Now she is an associated professor in the school of Management and Economics, Beijing Institute of Technology. She has ever been in Technology University of Berlin, Germany to do her Ph.D. research work from 2002 to 2003. In 2009 she has visited the Karlsruhe Institute of Technology, Germany for half a year.

Her current research interests include theory and method of information systems, e-commerce, semantic retrieval, intelligent information system etc.