

Research Article

The Collaborative Filtering Method Based on Social Information Fusion

Hao Wang , Yadi Song , Peng Mi, and Jianyong Duan 

Information College, North China University of Technology, Beijing, China

Correspondence should be addressed to Jianyong Duan; duanjy@hotmail.com

Received 22 January 2019; Accepted 18 March 2019; Published 2 April 2019

Academic Editor: Jean Jacques Loiseau

Copyright © 2019 Hao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the social network, similar users are assumed to prefer similar items, so searching the similar users of a target user plays an important role for most collaborative filtering methods. Existing collaborative filtering methods use user ratings of items to search for similar users. Nowadays, abundant social information is produced by the Internet, such as user profiles, social relationships, behaviors, interests, and so on. Only using user ratings of items is not sufficient to recommend wanted items and search for similar users. In this paper, we propose a new collaborative filtering method using social information fusion. Our method first uses social information fusion to search for similar users and then updates the user rating of items for recommendation using similar users. Experiments show that our method outperforms the existing methods based on user ratings of items and using social information fusion to search similar users is an available way for collaborative filtering methods of recommender systems.

1. Introduction

Importing social information to the collaborative filtering method is proved to be an available way to improve the performance of the recommender system and alleviate the data sparseness problem caused by the cold start [1–3]. Take, for examples, in microblog, the Chinese twitter, most existing collaborative filtering methods use user ratings of items (microblogs) to search for similar users. When the number of microblogs is much more than the number of microblogs which have been rated, the data sparseness problem will be caused. This problem makes searching similar users only relied on microblogs (items) inconvenient.

In order to address the above issues, we propose a new collaborative filtering method which uses social information fusion to search the similar users and then update user ratings of items using the similar users, which will alleviate the problem of data sparsity. Finally, we recommend the items based on the updated user ratings of items.

The remainder of this paper is organized as follows. In Section 2, we discuss the related work. Then we formulate the social information fusion in Section 3. The process of recommendation is given in Section 4. Finally, we show our

experimental results in Section 5 and draw a conclusion in Section 6.

2. Related Work

The user-based collaborative filtering recommendation algorithm assumes the interest between similar users (neighbors) is similar [4]. First, find the neighbor user of the target user, and then predict the target user's rating of the corresponding item according to the rating of the similar user, thereby completing the recommendation. To find neighbor users, the traditional method is to treat the user's rating of the item as a high-dimensional vector and use cosine similarity to find the user with high similarity as the target user's neighbor. This method can effectively use the feedback information of other users. However, in the actual environment, the number of items is far greater than the number of users, and the number of items that users can browse is limited, which results in the sparseness of the user-item rating matrix, which makes it difficult to find similar users. In the case of sparse data, a small number of user ratings is not enough to support larger items recommendation.

To alleviate this problem, literature [5] proposes a simple recommendation algorithm that fully exploits the similarity information among users and items and intrinsic structural information of the user-item matrix. This method constructs a new representation which preserves affinity and structure information in the user-item rating matrix and then performs recommendation task. Experimental results show the effectiveness of this method. Recently, a linear sparse and low-rank representation of the user-item matrix has been applied to produce Top-N recommendations. Literature [6] addresses the limitations of the nuclear norm by mimicking the behavior of the true rank function and improves the quality of recommendation substantially. Literature [7] proposes a method to reduce the item's space dimension based on singular value decomposition (SVD), but the reduction of dimension led to the loss of information at the same time, which makes the effect difficult to guarantee. Many scholars have begun to study the similarity between users from user attribute and social attribute. Literature [8] adopts the thought of deep search to find the trust relationship among users and then to regard people with the strong relationship of trust as similar users. Literature [9] presents Matchmaker, a collaborative filtering friend recommendation system based on personality matching, and gets a good result. Literature [10] studies the problem of link recommendation in weblogs and similar social networks and proposes an approach based on collaborative recommendation using the link structure of a social network and content-based recommendation using mutual declared interests. Literature [11] utilizes the user's interaction relationship to measure the user relationship strength. Literature [12] fuses the user's similarity of attributes and the user's interaction relationship for user similarity calculation. In addition, the analysis of user relationship is also often used for friends' recommendation [13].

Our study finds that every piece of social information reflects an aspect of the user. This paper proposes a novel method based on social information fusion instead of the method based on user ratings of items to search neighbors.

3. Fusion of Social Information

Based on the experimental data provided by KDD CUP 2012 track1, this paper divides a user's social information into personal description information (Profile), relationship information (Follow), behavior information (Action), and interest information (Interest). Therefore, the social information fusion is expressed as

$$\begin{aligned} & UserInformation(u) \\ &= Profile(u), Follow(u), Action(u), Interest(u) \end{aligned} \quad (1)$$

The similarity between the users is expressed as

$$\begin{aligned} sim(u_1, u_2) &= k_1 simProfile(u_1, u_2) \\ &+ k_2 simFollow(u_1, u_2) \end{aligned}$$

$$\begin{aligned} &+ k_3 simAction(u_1, u_2) \\ &+ k_4 simInterest(u_1, u_2), \\ &k_1 + k_2 + k_3 + k_4 = 1 \end{aligned} \quad (2)$$

The description of each piece of information and its similarity calculation are as follows.

3.1. Similarity of Personal Description Information. Personal description information includes gender (sex), age (age), and label (tag). The similarity of personal description information is defined as

$$\begin{aligned} simProfile(u_1, u_2) &= k_1 simSex(u_1, u_2) \\ &+ k_2 simAge(u_1, u_2) \\ &+ k_3 simTag(u_1, u_2), \\ &k_1 + k_2 + k_3 = 1 \end{aligned} \quad (3)$$

(1) $simSex(u_1, u_2)$. Gender is represented by numbers. Number 1 means male, Number 2 means female, and Number 0 means unknown. The similarity of gender is defined as follows:

$$\begin{aligned} & simSex(u_1, u_2) \\ &= \begin{cases} 1 & (u_1.Sex! = 0 \&\& u_2.Sex! = 0 \&\& u_1.Sex == u_2.Sex) \\ 0 & (u_1.Sex == 0 \parallel u_2.Sex == 0 \parallel u_1.Sex! = u_2.Sex) \end{cases} \end{aligned} \quad (4)$$

(2) $simAge(u_1, u_2)$. Age is expressed in year of birth. First, we transfer the user's birth year into an actual age. Second, consider that people's cognitive ability is changing along with age. For the older, the cognitive ability difference is bigger. For two people of younger and two people of older, even if these two groups have the same age difference, the cognitive ability difference between younger group will be smaller than that in older group in most situations. The similarity of age is defined as follows:

$$\begin{aligned} simAge(u_1, u_2) &= \frac{1}{2} \left(1 - \frac{|u_1.Age - u_2.Age|}{u_1.Age} + 1 \right. \\ &\quad \left. - \frac{|u_1.Age - u_2.Age|}{u_2.Age} \right) \end{aligned} \quad (5)$$

(3) $simTag(u_1, u_2)$. Tag is expressed in words. Due to the fact that different users have various tags, we use the proportion of the number of the two users' common tags and total tags as evaluation. The formula is as follows:

$$\begin{aligned} & simTag(u_1, u_2) \\ &= \frac{u_1.Tag \cap u_2.Tag}{u_1.Tag + u_2.Tag - u_1.Tag \cap u_2.Tag} \end{aligned} \quad (6)$$

3.2. *Similarity of Relationship Information.* Relationship information contains the focusing information (followee) and fans information (followers). Relationship information similarity is expressed as formula (7), and $k_1 + k_2 = 1$.

$$\begin{aligned} \text{simFollower}(u_1, u_2) &= k_1 * \text{simFollower}(u_1, u_2) + k_2 \\ &* \text{simFollower}(u_1, u_2) \end{aligned} \quad (7)$$

Focusing information and fans information are mutual; for example, user u_1 focuses on the user u_2 , the user u_2 is the user u_1 's focusing information, and user u_1 is u_2 's fans information at the same time. So, the calculation of focusing information and fans information is the same. Due to a large number of users in the data, high-dimensional cosine similarity calculation should be avoided, so we use the proportion of the two users' focusing/fans information and of their total as evaluation. The formula is as follows:

$$\begin{aligned} \text{simFollowee}(u_1, u_2) \\ = \frac{(u_1.\text{Followee} \cap u_2.\text{Followee})}{(u_1.\text{Followee} + u_2.\text{Followee} - u_1.\text{Followee} \cap u_2.\text{Followee})} \end{aligned} \quad (8)$$

3.3. *Similarity of Behavior Information.* Behavior information includes comment number (Comment), pointing frequency (At), and forwarding frequency (Retweet). Similarity of behavior information is expressed as follows, and $k_1 + k_2 + k_3 = 1$:

$$\begin{aligned} \text{simAction}(u_1, u_2) &= k_1 \text{simComment}(u_1, u_2) \\ &+ k_2 \text{simAt}(u_1, u_2) \\ &+ k_3 \text{simRetweet}(u_1, u_2) \end{aligned} \quad (9)$$

The above three types of data have the same forms, which led to the same evaluation methods. Taking comment number information, for example, creates a comment matrix based on the number of users. In comment matrix, each row stands for a user's number of comment given to other users, and each column stands for the number of comments of a user received. User u_1 's comments for other users are recorded vector $\overrightarrow{\text{Comment}}_{u_1}$. The similarity of two users is calculated using cosine similarity; the formula is as follows:

$$\begin{aligned} \text{simComment}(u_1, u_2) \\ = \frac{\overrightarrow{\text{Comment}}_{u_1} \cdot \overrightarrow{\text{Comment}}_{u_2}}{\|\overrightarrow{\text{Comment}}_{u_1}\| * \|\overrightarrow{\text{Comment}}_{u_2}\|} \end{aligned} \quad (10)$$

3.4. *Similarity of Interest Information.* Experimental data provide the keywords and the corresponding weights of keywords, both extracted from user's microblog comments, and classify all the keywords. According to this information, we could calculate the similarity of the user's interest information.

As is shown in Figure 1, keywords are included in one or more categories. After counting, the weight of every keyword

TABLE 1: The initial rating of the item.

	Item 1	Item 2	Item 3	Item 4	Item 5
User u	1	1	0	0	0
User v	0	0	1	0	0

could be reflected in the classification. Establish the interest classification matrix based on the number of classifications, and then every user could get a corresponding interest classification vector through calculating. The similarity of interest would be calculated using cosine similarity.

The calculation process is as follows.

Step 1. Set keywords set as $C = \{c_1, \dots, c_m\}$; set user-keywords weight as $U_C = \{\text{weight}_{c_1}, \dots, \text{weight}_{c_m}\}$; set classified collection as $V = \{v_1, \dots, v_m\}$. All data in above can be provided directly by the experimental data.

Step 2. Set the user-classification weight which we are going to calculate as $U_V = \{\text{weight}_{v_1}, \dots, \text{weight}_{v_m}\}$; the calculation formula of weights is $\text{weight}_{v_i} = (1/\|v_i\|) \sum_{c_j \in v_i} (\text{weight}_{c_j} / \sum_{c_k \in v_i} \text{weight}_{c_k})$. In this formula, $\|v_i\|$ means the keywords' number included in the classification of v_i .

Step 3. Regard $U_V = \{\text{weight}_{v_1}, \dots, \text{weight}_{v_m}\}$ as $\overrightarrow{\text{Interest}} = (\text{weight}_{v_1}, \dots, \text{weight}_{v_m})$.

Then, the similarity calculation of interest information between two users is as follows:

$$\text{simInterest}(u_1, u_2) = \frac{\overrightarrow{\text{Interest}}_{u_1} \cdot \overrightarrow{\text{Interest}}_{u_2}}{\|\overrightarrow{\text{Interest}}_{u_1}\| * \|\overrightarrow{\text{Interest}}_{u_2}\|} \quad (11)$$

4. The Process of Recommendation

User-based collaborative filtering recommendation creates a rating matrix based on user ratings of items and predicts the possible rating of the target user in the unrated item according to the rating matrix. The recommendation sequence is then generated based on the predicted rating. The sparseness of the rating matrix in the actual data is one of the key factors affecting the accuracy of the entire recommendation. The matrix filling technique [14] provides an idea in solving the matrix sparsity problem. For user's unrated content, early scholars set it to the average of other users' ratings of items [15], which did not fundamentally alleviate the sparseness of the rating matrix.

4.1. *The Introduction.* Based on the experimental data, the data provides a representation of whether the user likes the target item, each of which contains several keyword attributes. In the data, "1" means like, and "-1" means dislike. We use "1" or "-1" as the user's rating for the item and for the unrated item as "0".

As is shown in Figure 2, user u likes Items 1 and 2; user v likes Item 3. The initial rating is shown in Table 1.

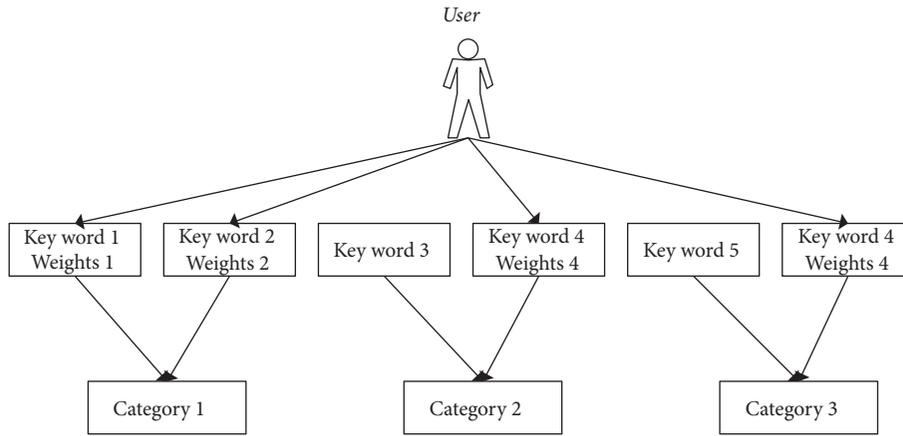


FIGURE 1: The relationship of user's interest.

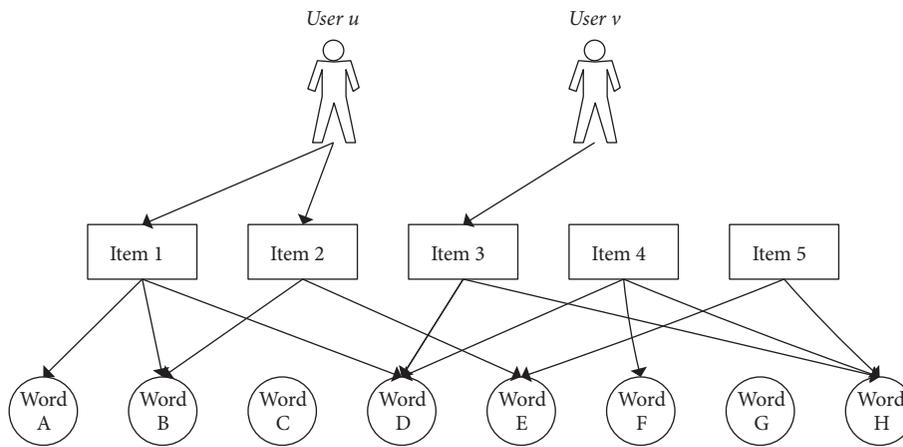


FIGURE 2: User ratings of items.

The data is sparse in the matrix, at the same time there are no common rating items between the two users. But there is a connection between the item they like. Taking Item 1 as an example, we divide the user u 's ratings for Item 1 into three equal parts according to the number of keywords included in Item 1 and think that word A, word B, and word D occupy 1/3 of the rating respectively. Excluding the relationship between Item 1 and word D, word D has a relationship with Item 3 and Item 4. At this time, the 1/3 rating of Item 1 occupied by the word D is equally divided into two and is reflected to the keyword D of Item 3 and Item 4; that is, the keyword D of the Item 3 and Item 4, respectively, occupies the 1/6 rating. Then process other items in this way. The processed ratings of the items are shown in Table 2.

This matrix filling process can pass the rating through the implicit relationship between the items to other related items, showing an implicit rating. At the same time, since some ratings of items are the same (1 or -1), personal preferences cannot be highlighted. Through this process, they can be distinguished.

4.2. *The Process.* Regard the sets of items as $Items = \{item_1, \dots, item_n\}$; regard the sets of words as $Words = \{word_1, \dots, word_m\}$.

Create an array, $SumEvaluation[n]$, to save the reflection rating.

Step 1. Select $item_i$ from $items$; the rating of $item_i$ is 1; extract $ItemWords = \{itemword_1, \dots, itemword_k\}$ from the set $words$; and in the formula, $word_j \in item_i$.

Step 2. Select $word_j$ from $Words$; find $TagetItems = \{tagetitem_1, \dots, tagetitem_x\}$ out; in the formula, $itemword_j \in tagetitem_z$ and $item_i \notin Tagetitems$.

Step 3. If $TagetItems$ is not empty, add the implicit rating $1/TagetItems.Count$ to $SumEvaluation[n]$ corresponding to each element in $TagetItems$.

Step 4. If there is any comment in $ItemWords$, continue from Step 2.

Step 5. If there is any comment in $Items$, continue from Step 1.

Step 6. Add original ratings to $SumEvaluation[n]$.

Finally, represent the values in $SumEvaluation[n]$ as user ratings of items. Put the new $SumEvaluation[n]$ together; then we get the new rating matrix.

TABLE 2: The processed rating of the item.

	Item 1	Item 2	Item 3	Item 4	Item 5
User u	3/2	4/3	1/6	1/6	1/2
User v	1/4	0	1	1/4	1/2

4.3. *The Recommendation.* According to the methods in Section 3, we get the neighboring user sets U . Next, process every user's rating by the calculation method above, then we get the new ratings. Therefore, the formula for user u 's prediction rating $P_{u,item_i}$ of $item_i$ is as follows:

$$P_{u,item_i} = \bar{R}_u + \frac{\sum_{v \in U} sim(u, v) * (R_{v,item_i} - \bar{R}_v)}{\sum_{v \in U} sim(u, v)} \quad (12)$$

\bar{R}_u presents the user u 's average rating of items, $R_{v,item_i}$ presents user v 's rating for $item_i$, and $sim(u, v)$ presents the similarity between user u and v . Finally, the first few items with high ratings in the predicted ratings are used as the recommended set.

5. Experiments

This section compares the ideas presented in this paper through two sets of experiments. The experimental content includes

- (1) the user's relationship information's effect to similar users;
- (2) the collaborative filtering method based on matrix filling algorithm.

5.1. *The Experimental Data.* The experimental data in this paper is derived from the data provided in the KDD CUP 2012 track1. The data is based on Tencent Weibo for a period of time, extracting all kinds of social information of users. These include user information (gender, age, number of microblogs, and tags), relationship information, behavior information (forwarding information, pointing information-@, and rating information), keyword information (keywords, weights, and classifications), and items accepted and rejected by users and their ratings. Since the data-related content is encrypted, it is inconvenient for data screening. We screened the data as follows.

Step 1. Randomly select 20 users whose age information is between 15 and 40 years from all users and put them in the queue.

Step 2. Dequeue a user from the queue, use the user as an experimental user, and put the friends he follows into the queue.

Step 3. Repeat the second step, when the number of experimental user groups reaches the target number.

After the progress in above is completed, we get 10001 experimental users. The related information includes 846077

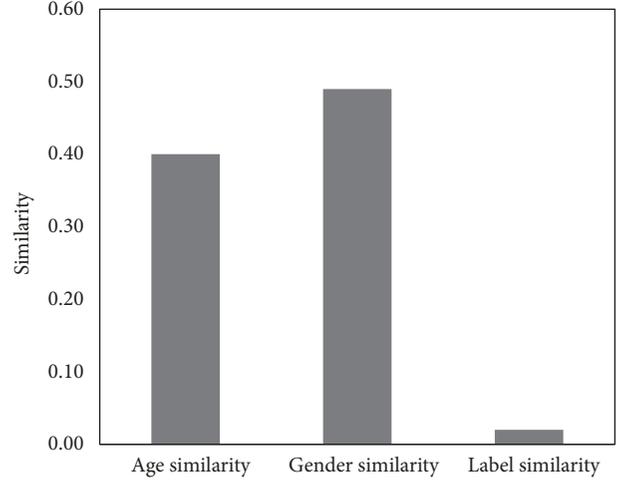


FIGURE 3: The comparison of age, gender, and label similarity.

pieces of item rating information, 821666 pieces of relationship information, 381208 pieces of behavior information, 16267430 pieces of keywords information, and 363 keywords' classification information.

5.2. *The Influence of the User's Social Information to Similar Users.* First, we divide the 10001 users into groups, each group consisting of two users, and then we get 50005000 similarity computing groups. For each group of users, according to the calculation process described in this paper, age similarity, gender similarity, label similarity, following similarity, fan similarity, forwarding similarity, pointing similarity, comment similarity, rating similarity, and interest similarity are calculated, respectively, as shown in Figures 3–6.

As can be seen from the data in Figure 3, the ratio of male to female in the experimental data is close to 1:1, and the age difference is moderate, but the label has a large difference. From the data in Figure 4, the fan information is more difficult to reflect the similarity between users than the information of interest. As is shown in Figure 5, the similarity of forwarding behavior in user behavior accounts for a large proportion. As is shown in Figure 6, the rating information can also be regarded as a kind of user interest, but compared with the keywords extracted in the experimental data and the interest extracted from the classification, it does not reflect the similarity of interest between users. This is due to the large number of microblogs, and the related microblog content cannot be browsed by the user, resulting in the inability to obtain the rating. Therefore, analyzing user interests according to their keywords and their related classifications in a user's limited page views can alleviate such problems.

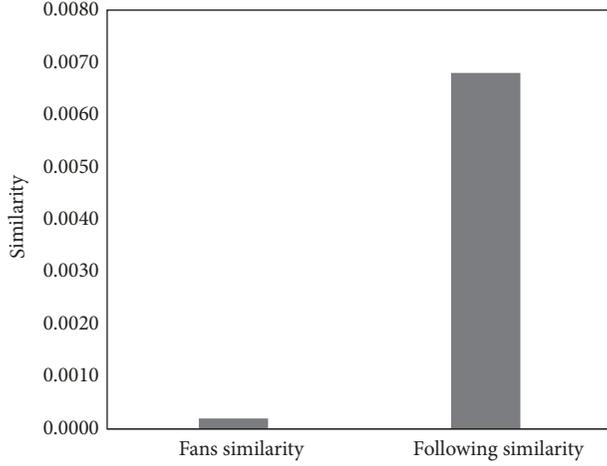


FIGURE 4: The comparison of fans and following similarity.

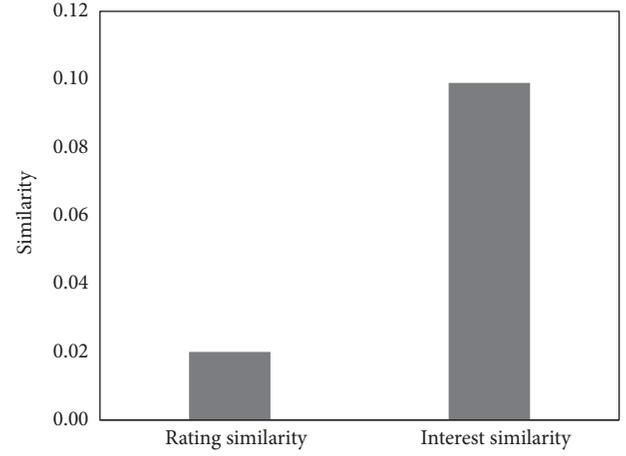


FIGURE 6: The comparison of rating and interest similarity.

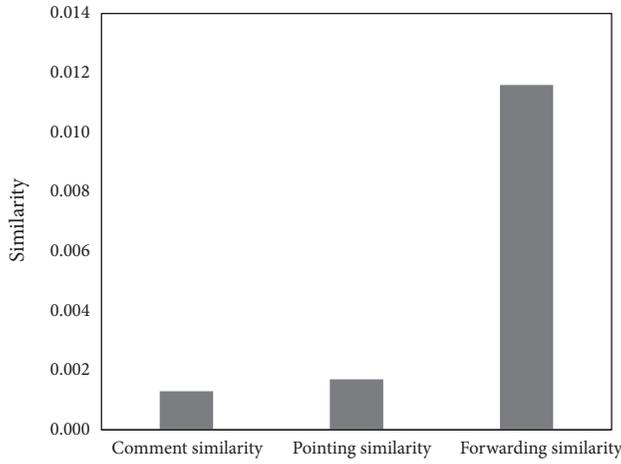


FIGURE 5: The comparison of comment, pointing, and forwarding similarity.

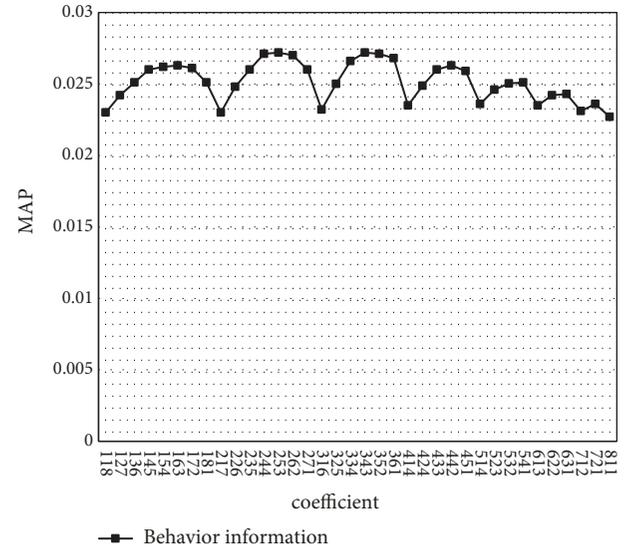


FIGURE 7: The MAP results' transformation of behavior information.

According to the similarity information calculated above, we use the personal description information, relationship information, behavior information, and interest information to perform the selection calculation of similar users and then find the fusion parameters in each piece of information. The evaluation criteria of similar users are based on the user's own followers. The description is as follows: select one user u from the experimental users and calculate the similarity for the other users. The descending sequence of similarity values is taken as the user u 's similar user sequence. Use the user u 's own followers as the correct collection. The similar user sequence is valued from 5 to 40 according to Top-N, and combined with the user set, the map (Mean Average Precision) calculation is performed [16], and the formula is as follows:

$$MAP@N = \frac{1}{|U|} \sum_{u \in U} \left(\frac{1}{|F(u, N)|} \sum_{i=1}^{|F(u, N)|} \frac{i}{S(u, N)_{\text{-rank}(i)}} \right) \quad (13)$$

In this formula, $|U|$ represents the number of experimental user sets; $|F(u, N)|$ represents the number of elements included in the correct recommendation set of user u when the recommendation number is N . $S(u, N)$ represents the first N similar user sets. $S(u, N)_{\text{-rank}(i)}$ represents the ranking of the i -th correct content in the similar user sets.

Taking behavior information as an example, according to the fusion calculation methods, $simAction(u_1, u_2)$ described in this paper that the coefficients of the three kinds of data included in the behavior information are valued in the range of 0.1 to 0.8, ensuring that each coefficient is greater than 0 and the sum of the three coefficients is 1. When Top-N takes 40, the average map value image of all experimental users is shown in Figure 7.

The horizontal axis represents the coefficient. From top to bottom, they are the coefficient of point information, forwarding information, and comment information. As shown in the graph, finding similar users based on behavior information is mainly affected by the similarity of the forwarding

TABLE 3: The coefficients.

The formulas	Parameter
The fusion formula of personal description information	the coefficient of label information: 0.5; the coefficient of age information: 0.25; the coefficient of gender information: 0.25
The fusion formula of relationship information	the coefficient of following information: 0.5; the coefficient of fans information: 0.5
The fusion formula of behavior information	the coefficient of comments information: 0.3; the coefficient of pointing information: 0.2; the coefficient of forwarding information: 0.5

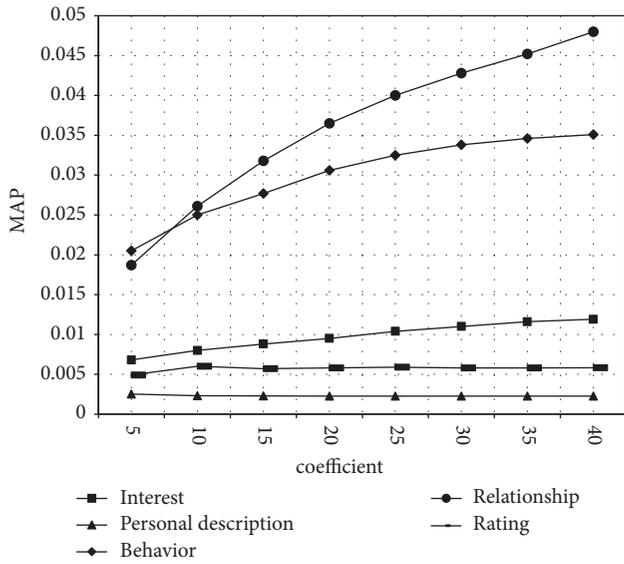


FIGURE 8: The MAP results of each attribute.

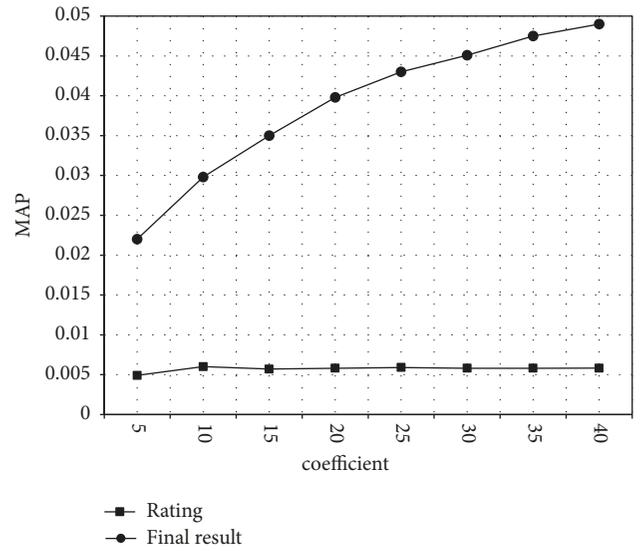


FIGURE 9: The comparison of final experiment result.

information. Each peak is generated when the forwarding information coefficient has a larger value, and each trough is generated when the forwarding information coefficient is the smallest.

We will get the maximum value of the map when the value on X-axis is “253”. So the formula of convergence computation is expressed as

$$\begin{aligned}
 simAction(u_1, u_2) = & 0.3 * simCommon(u_1, u_2) \\
 & + 0.2 * simAt(u_1, u_2) + 0.5 \quad (14) \\
 & * simRetweet(u_1, u_2)
 \end{aligned}$$

In the same way, we give the weight of personal description information and relationship information, as shown in Table 3.

Next, we conduct experiments to find similar users according to the combined personal description information, relationship information, behavior information, and interest information and rating information. In the experiment, the Top-N values were taken from 5 to 40, and the average map value of the experimental users is shown in Figure 8.

As can be seen from the figure, with the growth of Top-N, it is better to calculate similar users according to social

information than other methods. However, the map value calculated based on the rating information is low, and it is not suitable to find similar users based only on the content. The behavior information is better when the value of Top-N is lower.

The experimental data can show that, due to the sparse user ratings, implicit ratings are not enough to give reasonable similar users. Similar users based on relationship information and behavior information are better than rating information. According to the idea of fusion and through a large amount of data calculation, a similar user calculation formula $sim(u_1, u_2)$ is obtained, in which the coefficient of behavior information, relationship information, interest information coefficient, and personal description information are 0.36, 0.49, 0.11, and 0.44, respectively. Then compare with the traditional method of finding neighbor users according to the rating. The comparison result is shown in Figure 9, where Top-N ranges from 5 to 40.

5.3. The Collaborative Filtering Recommendation Experiment Based on the User Similarity Degree. The experimental data provided a record of 845,727 use ratings of items involving 3,775 related items. According to the method of finding

TABLE 4: The proportion of rating matrix sparsity.

	The proportion of nonzero	
	Before processed	After processed
The original rating matrix	0.044	0.397
The positive ratings rating matrix	0.018	0.174

neighbor users in Section 5.2, we selected 40 neighboring users for every user, and each user with the corresponding neighboring users comprised the evaluation matrix of the item. According to the method of processing matrix sparsity in Section 4, we process the rating matrix. The result is shown in Table 4.

According to the processed rating matrix, we repredict the current user's ratings of the item according to the formula in Section 4.3. We use the recall rate (R) and precision (P) and F1 as the basis for evaluation. The recall rate is the ratio of the correct recommendation quantity to the theoretical maximum recommendation item quantity; the precision rate is the ratio of the correct recommendation quantity to the recommendation total number; the F1 calculation can be balanced between the recall rate and the precision rate. The formulas are as follows:

$$R = \frac{|hits|}{|testset|} P = \frac{|hits|}{|recset|} F1 = \frac{2 * P * R}{P + R} \quad (15)$$

$|hits|$ indicates the number of items whose original rating of the user is "1" in the recommendation set. $|testset|$ indicates the number of all users whose original rating is "1". $|recset|$ indicates the number of items in the recommendation set. Then we sort the user's predicted rating results and select the recommendation set according to the Top-k. Three cases were selected for comparison in the experiment.

(1) Recommending directly without matrix sparsity processing.

(2) Recommending after matrix sparsity processing.

(3) Only considering the positive evaluation of the user, getting the rating matrix, and then performing the recommendation after matrix sparsity processing. Figure 10 shows the variation of the F1 value between 5 and 45 for the Top-k value.

According to the experimental results, we can get the following:

(1) The recommendation effect after sparse processing of the rating matrix is better than the recommendation effect of not processing the rating matrix. When the value of Top-k is 5, the two are basically the same. As the Top-k grows, the difference is obvious. When the value of Top-k is between 30 and 45, the difference between the two is relatively obvious. According to the result, when the matrix sparsity processing is not performed, the item with more original rating of "1" cannot be selected as the recommendation result.

(2) Negative ratings have a certain influence on the recommendation effect. The rating includes a positive rating with a value greater than zero and a negative rating with a value less than zero, with a value of zero indicating that the user has not rated. In the data, we found that the ratio

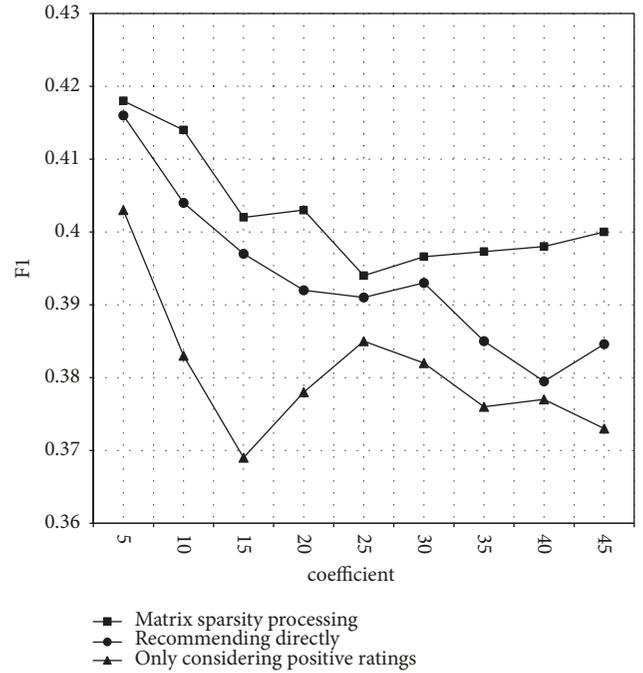


FIGURE 10: The experimental results of item recommendation.

of the negative rating is more important than the positive rating. After the rating vector processing, the values of more ratings are expressed as negative numbers. In order to reflect the impact of negative ratings, we removed the negative ratings, only considering the impact of positive ratings and conducting recommendation experiments. As can be seen from Figure 9, the experimental results considering both positive and negative ratings are better than those considering only negative ratings.

6. Conclusion

In this paper, based on the data set of KDD CUP 2012 track1, we study the users' social information and propose a new collaborative filtering method based on social information fusion. Experiments show that our new method outperforms the method based on user ratings of items and the social information fusion is an available feature for recommender system. In the future, we will study how to extract more effective social information fusion features using the deep learning method.

Data Availability

The data is a public dataset and everyone can download from <https://www.kaggle.com/c/kddcup2012-track1>. In Section 3, we provide the data name.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61672040) and the North China University of Technology Startup Fund.

References

- [1] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, "Joint social and content recommendation for user-generated videos in online social network," *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 698–709, 2013.
- [2] L. Quijano-Sanchez, J. A. Recio-Garcia, B. Diaz-Agudo, and G. Jimenez-Diaz, "Social factors in group recommender systems," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 1, p. 8, 2013.
- [3] Valdis Krebs, "Social network analysis, a brief introduction," 2008. <http://www.orgnet.com/sna.html>.
- [4] A. Bellogn, I. Cantador, F. Dez, P. Castells, and E. Chavarriaga, "An empirical comparison of social, collaborative filtering, and hybrid recommenders," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 1, p. 14, 2013.
- [5] Z. Kang, C. Peng, M. Yang, and Q. Cheng, "Top-N recommendation on graphs," in *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016*, pp. 2101–2106, USA, October 2016.
- [6] K. Zhao and C. Qiang, "Top-n recommendation with novel rank approximation," in *Proceedings of the 2016 SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics, 2016.
- [7] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of Dimensionality reduction in recommender system - a case study," Tech. Rep., Minnesota University Minneapolis Department of Computer Science, 2000.
- [8] M. Jamali and M. Ester, "TrustWalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, pp. 397–405, France, July 2009.
- [9] L. Bian and H. Holtzman, "Online friend recommendation through personality matching and collaborative filtering," in *Proceedings of the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, pp. 230–235, 2011.
- [10] W. H. Hsu, A. L. King, M. S. R. Paradesi, T. Pydimarri, and T. Weninger, "Collaborative and structural recommendation of friends using weblog-based social network analysis," in *Proceedings of the AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, vol. 6, pp. 55–60, USA, March 2006.
- [11] I. Kahanda and J. Neville, "Using transactional information to predict link strength in online social networks," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 9, pp. 74–81, 2009.
- [12] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks," in *Proceedings of the 19th International World Wide Web Conference, WWW2010*, pp. 981–990, USA, April 2010.
- [13] J. Hannon, K. McCarthy, and B. Smyth, "Finding useful users on twitter: twittomender the followee recommender," in *Proceedings of the 33rd European Conference on Information Retrieval (ECIR-11)*, DCU, Springer, Dublin, Ireland, April 2011.
- [14] Z. Kang, C. Peng, and Q. Cheng, "Top-N recommender system via matrix completion," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence, AAAI 2016*, pp. 179–185, USA, February 2016.
- [15] A.-L. Deng, Y.-Y. Zhu, and B.-L. Shi, "Collaborative filtering recommendation algorithm based on item rating prediction," *Journal of Software*, vol. 14, no. 9, pp. 1621–1628, 2003.
- [16] S. Rendle, "Social network and click-through prediction with factorization machines," in *Proceedings of the KDD-Cup Workshop*, 2012.

