

A hybrid recommendation algorithm based on social networks

Jing Gong

College of mathematics &
physics of Nanjing
University of Posts and
Telecommunications
Key Laboratory of
Broadband Wireless
Communication and Sensor
Network Technology
Nanjing, China
gj_jing@139.com

MeiLing Gao

Key Laboratory of
Broadband Wireless
Communication and Sensor
Network Technology
Nanjing, China
gml814888@163.com

Bixiao Xu

College of mathematics &
physics of Nanjing
University of Posts and
Telecommunications
Nanjing, China
xubixiao@139.com

Wenjun Wang, Zhixin Sun

Key Laboratory of
Broadband Wireless
Communication and Sensor
Network Technology
Nanjing, China
wangwenjun @163.com
sunzx@njupt.edu.cn

Abstract—In the light of the problem of collaborative recommendation and content-based cold start, this paper proposes a hybrid recommendation system based on social network. The method is based on the user social relations network. According to the social behavior of user, by establishing the model of social network users, it puts forward the user similarity measure. Then it takes random walk algorithm as a basis and selects out N users who have the highest similarity with the users' interest. The test results show that this method can obtain better recommendation effect and customer satisfaction.

Keywords- Social network; Recommendation algorithm ; Collaborative recommendation

I. INTRODUCTION

Social networking has a broad user community. Its swiftness and the way of open spreading information are extremely attractive for commodity marketing, so it has great commercial value to introduce goods show part into social networking sites. However, the chaotic, uneven and untargeted promotion of goods is extremely easy to cause the user antipathy, which makes the social networking sites and businesses suffer losses. Therefore, the establishment of user personalized demands and the recommendation of the goods which are interested and trusted by users both provide a good display platform for merchants, and provide users with a friendly recommendation, which makes the social networking sites, merchants, users tripartite altogether win.

II. RESEARCH ON CORRELATIVE TECHNOLOGY

In today's social network recommendation algorithm, the most common methods are collaborative filtering recommendation algorithm and recommendation algorithm based on content.

A. Collaborative Filtering Recommendation Algorithm

Collaborative filtering recommendation method [1] was put forward by Goldberg and others in 1992. Over the past 15 years,

it has been widely studied and applied. The basic idea of collaborative filtering recommendation was refer to the method of people's daily purchase. That is to say, when people buy goods, they generally tend to get advice from friends who have similar interests with them. Or users like certain goods. Then, when some goods are similar to the users' favourite type and have a high evaluation, the purchase probability of this type of commodity will be higher.

The literature [2-8] mainly introduced the method of using collaborative filtering in social networks to calculate the recommendation. Sun GF[2] et al. proposed a method for modeling the temporal behavior of a user (product), but there was no detailed introduction on recommendation model. Huang Chuangguang [3], et al proposed a kind of uncertain neighbor collaborative filtering recommendation algorithm which does not consider the interaction between users and changes of the attention to goods. In literature [4], the author recommended by excavating the user preference. Although it considered personalized features, there were still some problems on data sparseness in the data processing method.

Although collaborative filtering recommendation has been widely researched and largely applied, there are also some problems about cold start, data sparseness, etc. Especially the collaborative filtering data source is the evaluation information of user item. But for providing commodity recommendation service in social networking sites, it is not applicable.

B. Recommendation Algorithm Based On Content

Recommendation based on content originally derived from research in the field of information retrieval. Its main idea is to recommend goods whose features are similar with the things users ever bought to users. Namely, according to the historical behavior data of the target users, construct user preference document and choose high matching degree on feature with preference document from the recommended project to recommend to the user. There are two core works: one is feature description of recommendation project and user interest, the

Identify applicable sponsor/s here. (*sponsors*)

second is to calculate similarity between the content features of the target project and the interest features of users [9,10].

Literature [11-14] proposed that on the basis of the user clustering with the same interest ,the algorithm firstly through the user access log information excavate the similar content recommendation items. Then according to user interests excavate new content to recommend. Synonymy and polysemy of the keywords proposed by Literature [12,13] often lead to inaccurate calculation of document similarity, which is often difficult to balance in the recommendation of social network.

Recommendation method based on the content is easy to realize and comprehend. Besides, it has visualized recommendation results and doesn't need professional knowledge. But there are some difficulties, such as feature extraction ,the discovery of the users' potential new interest etc..Especially the problems of feature extraction ,for video, images, music and other complex unstructured object, there is no effective solution.Even text class project, feature extraction can only reflect part of its content, which may affect the recommendation satisfaction.Therefore ,recommended method based on content can not be fully competent in mobile social networking sites to provide users with product recommendation service, but it can be used as an effective adjunct method for recommendation.

In summary, there are some deficiencies in collaborative filtering and recommendation based on content and plus a lot of e-commerce site lacks personalized data of users, so it is difficult to give satisfactory recommendation.To solve these problems, this paper proposes a hybrid recommendation algorithm based on social networks. This algorithm based on the users social relations network, excavates N users who have the highest similarity with the user interest .At the same time, considering user message of the social network may contain user demand for commodities, so excavating the user message and applying the mining information to the recommendation can obtain better recommendation quality and user satisfaction.

III. SOCIAL NETWORK STRUCTURE AND USER SIMILARITY MEASURE

A. A Social Network Structure Model Based On The Degree Of Friend Attention

Recommendation algorithm of this paper is based on the social network of community.Users don't have certain attention or friends relationship in this the social network.But different users can belong to the same community.The same users also can join different community.The users in the same community said that they may have a similar interest.The group of douban is a typical example. Because the undirected graph of bidirectional established social network can convert undirected edge to two directed edges.So this paper is based on the social network of unidirectional attention.

In social networks, users establish social network with other users by concerning and communicate with each other by mention, forwarding, comments and other social interaction behaviors.As shown in Figure 1, the graph $G(V, E, w)$ defines a social network.V is the vertex set. Each vertex represents a

user.E is the edge set. If user V_a and V_b user have a social network relationship, then there is an edge $e(V_a, V_b)$ to connect the two users. $w(V_a, V_b)$ defines the weight of edges.

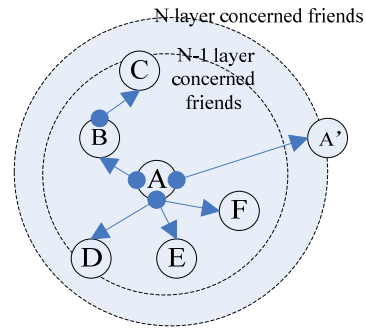


Figure 1. Social network relationship model

Definition : N layer concerned friends

A layer concerned friends: when user A focuses on user F, it will form a edge relationship (as Figure 1).So it defines F as a direct concerned friends or a layer concerned friends.

Two layer concerned friends: when user A focuses on user B,user B focuses on user C and user A does not directly focus on user C, it defines user C as two layer concerned friends of user A (as Figure 1).

N layer concerned friends,it defined user v as the N layer concerned friends of user A in this paper:Suppose a user v is n-1 layer concerned friends of user A,user v concerns the user u and user A along with its n-2 layer concerned friends do not directly concern the user u ,which defines user u as the N layer concerned friends of user A(as shown in Figure 1).

B. User Similarity Measure Standard in Social Networks

From the social network relational model (Figure 1), the weights of the edges should be able to characterize the similarity between users, including familiarity / trust degree and interest similarity.The studies of literature [15,16] have confirmed the existence of positive correlation between user interests similarity and user trust degree .However, with the rise of online social networks, in addition to similar interests ,we are often affected by the recommendations of friends when we are shopping in our daily life.So how to measure the similarity of interest / trust among users, it needs to determine a measure in recommendation algorithm .

So this paper by data mining, extracting feature vectors and using user features calculates the user favorability .Set the characteristics sets of user u_i and user u_j as F_i and F_j . Then defines the goodwill degree of user u_i to the user u_j as:

$$favor(u_i, u_j) = \cos \frac{F_i \cdot F_j}{|F_i| |F_j|}$$

(Formula 1)

However, the similarity of interest / trust between users is variable. If the user u and the user v are engaged in similar work and often exchange,the trust degree among them will improve. And interest similarity / trust degree among users is

time-sensitive. U in the past trusted v does not mean that u will always has a higher degree of trust in v. The performance of users and other related information makes u reevaluate the trust to v.

But no matter how the user interest similarity /trust degree fluctuates, the interaction between them will always be shown on their interaction of social networks .On social networks if user u often operate (comment, forwarding, etc.)on the user v message or user u in the messages always don't forget to @ user v,it often not only reflects that the user u is goodwill and trust in user v,but also reflects that user u is interested in user v. Therefore, this paper proposes to use the interaction degree of the user u to user v of recent period of time T as a measure of the similarity of the user u to user v.

To define the interaction degree of user u to user v is :

$$Interaction(u,v) = \frac{R_{uv} + T_{uv} + A_{uv}}{M_v + A_u}$$

(Formula 2)

M_v represents the number of user v's published news(or called weibo) within time T. R_w represents the number of user u's replies to user v within time T. T_w represents the number of user u's forward to user v. A_u represents the number of messages with @ sent by the user u within the time T. A_w represents the number of messages which are sent by the user u with @ user v within the time T and there is

$$\sum_{v=1}^n A_{uv} = A_u$$

But considering each user social network behavior habit is different. Some users keen to interact with other users and some users are more introverted whose interaction frequency is not high, but it does not mean that these users are not very trusted / interested in other users. Therefore, we consider the total interaction rate of the users within the time T.

Define the overall interaction rate of user u as:

$$Interaction(u) = \frac{\sum_{i \in out(u)} (R_{ui} + T_{ui}) + A_u}{\sum_{i \in out(u)} M_i + M_u}$$

(Formula 3)

$\sum_{i \in out(u)} M_i$ represents that user u node directly points to all the user nodes, that is to say, it is the number of messages delivered by all users concerned by user u.

$\sum_{i \in out(u)} (R_{ui} + T_{ui})$ represents that the number of replies and retransmissions to all users concerned by user u.

At last, this paper defines the edge weight of user u pointing to user v as:

$$e_{ui} = \alpha + \beta \cdot \frac{Interaction(u,v) - Interaction(u)}{Interaction(u)}$$

(Formula 4)

α represents foundation trust value, that is, give the basis of trust to all friends concerned by user u, besides, there is $\alpha + \beta = 1$.

IV. USER RECOMMENDATION ALGORITHM BASED ON RANDOM WALK ALGORITHM

Considering that in social network different users may extremely differ in the number of their concerned friends, if we get the Top-N users in similarity only from the direct concerned friends, it is difficult to select a suitable parameter N. Besides, for those users with less concerned friends, it may exist the problem of less available information and poor effect recommendation during the subsequent recommendation. On the other hand, if the similarity of user A and user B is very high, while similarity of user B and user C is also very high, if only considering the direct concerned friends of user A, it will generate the transitivity loss of neighbor relation. Therefore, it is not advisable to get the Top-N users in similarity only from the direct concerned friends. We should use the information of multilayer concerned friends to smooth the gap of the number of friends concerned by different users and extensionally dig the users with high similarity.

This article uses the method of gradual increase to determine the level of the concerned friends involved in the calculation, and restarts the random walk algorithm to use in this environment. Then, we calculate the similarity between multilayer users and the target users. Finally, we can get Top-N users in similarity.

A. Recommendation Based on Random Walk Algorithm

Random walk algorithm is a representative graph ranking algorithm, and it can calculate the correlation between any two nodes in the graph. Its operation principle is: start from the source point, randomly walk in the figure, decide to walk down or end of the walk and back to the source point according to the probability of α when it arrives at each node. If it continues to walk, it will select randomly the next walk node according to the probability distribution from the node which the current node points to. So after several random walks, each node will converge to a fixed value by the passing probability, which can be also called the correlation degree of the point and the source point.

The Markov nature of the random process can be expressed as:

$$P_r(X_{u,k+1} = i | X_{u,k})$$

(Formula 5)

The transition probability $P_{i,j} = P_r(X_{u,k+1} = j | X_{u,k} = i)$ indicates that the probability of the transfer of moment k+1 to the state j when the moment k is in the condition of the state i.

The restart random walk algorithm is a node ranking algorithm based on structure correlation. Compared with other similarity measurement methods, restart random walk algorithm has the advantages of capturing multi lateral relations between the random walk method capture graph global structure as well as the vertex is a good metric graph two vertices in a similarity method. In this paper, the similarity of

multilayer friends in social networks friends concerned graph is calculated by using the restart random walk algorithm.

The mathematical expression of RWR is as follows:

$$R_{t+1} = (1 - \delta)S \cdot R_t + \delta q$$

(Formula 6)

R_t , R_{t+1} and q are column vectors. R_t represents the probability distribution in step t . $R_{(t,i)}$ represents the probability that the step t arrives at vertex i . Column vector q is the restart vector and the unit vector whose the first dimensional is one, which represents the initial state. q_i represents the probability that vertex i is the starting vertex. S is the transition probability matrix. The element S_{ij} in the matrix represents the probability of transition from the current vertex i to the vertex j at the next step. When the vertex skips, the probability of directly returning to the starting point is σ . the probability of jumping to the immediate neighbor vertex which is the most similar to the current vertex is $1 - \sigma$. We use the formula (5) to calculate the probability distribution of the vertices in the walking process.

This probability distribution characterizes the probability that every vertex is visited in the graph. Use this probability distribution as the input of the next walk, and after repeating the iteration R is convergent. The probability of the graph reaches a stable distribution, and then it will be never changed by the next iterative. Therefore, the convergent probability is regarded as the similarity between concerned friends and target users. Sort similarity and get Top-N users in similarity.

B. User Recommendation Algorithm Based on Random Walk Algorithm

Based on the aforementioned user similarity measure, we get the goods concerned by other T-10 users who are worthy of the trust of users. So according to the random walk algorithm, the detailed steps to mine the Top-N users are as follows:

Step 1: Set the number of preliminary election concerned friends as M .

Step 2: Establish direct concerned friends for the target users who need recommendation. Calculate the similarity and do statistics to obtain the number of the direct concerned friends marked as N_1 .

Step 3: Compare N_1 and M , if $N_1 \geq M$, then skip this step. If $N_1 < M$, according to similarity ranking of a layer concerned friends from high to low to establish two layer concerned friends of target users and calculator the similarity of a layer concerned friends who have the direct connection with two layer concerned friends. Do statistics to obtain the number of the second layer concerned friends, marked as N_2 . Compare M with $N_1 + N_2$, if $N_1 + N_2 \geq M$, then end this step, or by analogy to build the next layer of concerned friends (as shown in figure 2).

Step 4: According to the primary concerned friends figure of established target user, use the restart random walk algorithm to calculate the similarity among other users in the figure of

target audience and primary concerned friends, until choose the N users which the highest similarity.

V. PERFORMANCE TESTING OF HYBRID RECOMMENDATION SYSTEMS

A. Experimental Data and Evaluation Index

There are two general categories for the evaluation of the recommendation system. One is to measure the accuracy of predicted score, in other words, it is to measure the degrees of deviation between the recommended prediction score and users' actual score for the recommendation. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are the common measure methods. The other is to measure the accuracy of the recommendation classification, that is to measure whether the recommended results conform with users' preferences. The common measure methods are recall and precision.

Considering the intrinsic property and practical applications of the hybrid recommender system based on the social network, this experiment takes precision as the index of measuring the accuracy of recommendation. Precision is defined as

$$precision = \frac{1}{T} \sum_u \frac{N_p}{L}$$

T is the number of test samples. L is the length of recommended list. N_p is the number of users' favor in the recommendation list.

B. The Recommendation Realization of Top-10 User of Interest Similarity

The test data sets of hybrid recommendation system based on social network includes 659 normal users' attention relationship in Youxin system, published microblogging, interaction data (based on the filtration system). Besides, it also includes user commodity information of 178 businesses and ordinary users' purchase, evaluation, collection and other datas within three months.

In order to analyze the performance of the recommendation algorithm, various parameters of the algorithm are set up firstly.

TABLE I. PARAMETER SETTINGS FOR RECOMMENDATION ALGORITHM

parameter name	parameter value	explanation
M	50	the number of preselected concern friends
N	10	N users with the highest similarity in the selection
K	10	the number of recommending commodities K
α	0.5	α represents base trust value
β	0.5	Variable trust value
λ	0.3	percentage of similarity in the recommendations
δ	0.7	the proportion of commodity evaluation in the recommendation

This paper establishes their own direct concerned friends table for the 659 normal users in Youxin system and calculates

the direct similarity. Table 2 shows direct concerned friends and similarity of a part of the 659 users, such as the directly similarity of U135 who is the direct friends of user U12 is 0.86.

TABLE II. THE TABLE ON FRIENDS DIRECTLY CONCERNED BY PARTIAL USERS AND SIMILARITY

User	Direct concerned friends
U12	(U3,0.41), (U9,0.87), (U13,0.77), (U14,0.80), (U15,0.74), (U21,0.31), (U23,0.61), (U55,0.65), (U58,0.31), (U73,0.16), (U97,0.58), (U112,0.68), (U132,0.79), (U133,0.23), (U135,0.86)
U97	(U12,0.67), (U21,0.71), (U59,0.67), (U73,0.79), (U74,0.73), (U75,0.74), (U83,0.72), (U93,0.81), (U94,0.65), (U96,0.57), (U99,0.47), (U112,0.91), (U113,0.68), (U126,0.35), (U134,0.21), (U135,0.62), (U136,0.43), (U147,0.66), (U168,0.45), (U242,0.48), (U257,0.72)
U243	(U13,0.77), (U14,0.80), (U15,0.74), (U21,0.31), (U96,0.57), (U99,0.47), (U112,0.91), (U113,0.68), (U126,0.35), (U134,0.21), (U135,0.62), (U136,0.43), (U147,0.66), (U168,0.45), (U242,0.48), (U257,0.72)
U465	(U14,0.80), (U15,0.74), (U21,0.31), (U126,0.35), (U134,0.21), (U135,0.62), (U168,0.45), (U242,0.48), (U257,0.72), (U279,0.79), (U280,0.73), (U299,0.89), (U369,0.83), (U385,0.63), (U414,0.62), (U415,0.85), (U435,0.25), (U489,0.42), (U521,0.51), (U533,0.79)

Afterwards, according to the concerned situation and the direct similarity, use the incremental method to establish the relation web. Take the user U12 as the example:

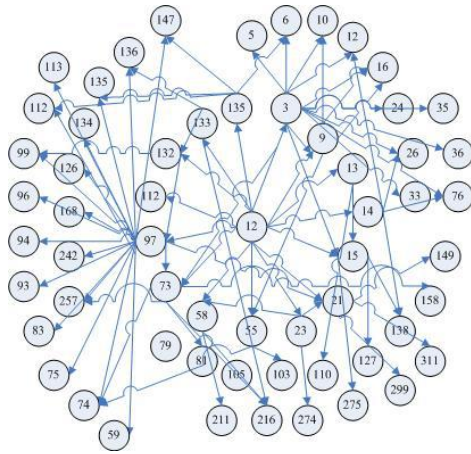


Figure 2. The concerned relation web of user 12

After that, a recommendation method based on random walk algorithm is used to calculate the Top-10 users in interest similarity of each target user. Partial results are presented in Table 3. For example, the Top-10 users in interest similarity

of user U12 is: U9, U 76, U14, U77, U135, U132, U15, U23, U133, U35.

TABLE III. THE TOP-10 USERS IN INTEREST SIMILARITY OF PARTIAL FRIENDS

User	Top-10 users in interest similarity
U12	U9, U 76, U14, U77, U135, U132, U15, U23, U133, U35
U97	U112, U73, U83, U45, U257, U75, U94, U88, U35, U79
U243	U112, U93, U15, U147, U257, U299, U113, U135, U235, U231
U465	U14, U369, U147, U257, U299, U385, U297, U379, U421

According to the users' purchase, evaluation, collection and other datas within three months, calculate the overall evaluation of goods and use method 2 to calculate the recommended degree of goods and finally feedback the results to the system user.

C. Comparison of Several Recommendation Algorithms

In order to validate the pros and cons between the recommendation algorithm proposed by this paper and the existing recommendation algorithm, this paper in the same data sets respectively validate the recommendation accuracy and consumed time between hybrid recommendation algorithm proposed by this text and recommendation algorithm based on content, collaborative filtering recommendation algorithm, such as table 4 shows.

TABLE IV. ACCURACY COMPARISON OF THREE RECOMMENDATION ALGORITHMS

Node format	200	600	1000	2000	3000
Recommendation algorithm based on collaborative filtering	77%	82%	86%	85%	78%
Recommendation algorithm based on content	75%	80%	85%	87%	80%
The hybrid recommendation algorithm in this paper	80%	85%	88%	88%	88%

As can be seen from the table, the accuracy of collaborative filtering recommendation algorithm and recommendation algorithm based on content are gradually increased at first with the increase of the number of nodes, but when the number of nodes are more than 2000, due to the excessive computation of content, instead of the generation of the decline in accuracy. Hybrid recommendation algorithm in this paper combines the advantages of the first two and considers the needs of social networks for recommendation, so accuracy almost remains stable and efficient when nodes are gradually increased.

V. CONCLUSION

This paper proposes a hybrid recommendation system of social network, based on the users social relations network, according to the social behavior of the user defines direct similarity between users, and calculates the similarity between indirect concerned friends by random walk algorithm, and then digs out the N users of highest similarity with recommended

users . Afterwards, by defining the overall value of goods directly manipulated by users, calculate recommended degree of target users with recommend commodities to recommend. The experimental data shows that the method can obtain a better recommendation result.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (NSFC) under grant No. 60973140 and 61170276, Jiangsu Provincial Natural Science Foundation of China under grant No.BK2009425, the Key University Science Research Project of Jiangsu Province under grant No.12KJA520003, and the Science and Technology Innovation Training Program (201410293023Z).

REFERENCES

- [1] Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry[J].Communications of the ACM. 1992,35 (12):61-70.
- [2] Sun GF, Wu L, Liu Q, Zhu C, Chen EH. Recommendations based on collaborative filtering by exploiting sequential behaviors. Ruan Jian Xue Bao/Journal of Software, 2013,24(11):2721 - 2733 (in Chinese).
- [3] HUANG Chuang-Guang,YIN Jian,WANG Jing,LIU Yu-Bao,WANG Jia-Hai. Uncertain Neighbors' Collaborative Filtering Recommendation Algorithm. CHINESE JOURNAL OF COMPUTERS,2010,33(8):1369-1377.
- [4] Yongli Ren ; Gang Li ; Wanlei Zhou .Learning User Preference Patterns for Top-N Recommendations. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on Volume: 1 ,Publication Year: 2012 , Page(s): 137 - 144
- [5] Guo L, Ma J, Chen ZM, Jiang HR. Learning to recommend with social relation ensemble. In: Proc. of the ACM Int' l Conf. on Information and Knowledge Management. ACM Press, 2012. 2599-2602.
- [6] Wu L, Chen EH, Liu Q, Xu LL, Bao TF, Zhang L. Leveraging tagging for neighborhood-aware probabilistic matrix factorization. In: Proc. of the ACM Int' l Conf. on Information and Knowledge Management. ACM Press, 2012. 1854-1858.
- [7] Liu Q, Chen EH, Xiong H, Ding CHQ, Chen J. Enhancing collaborative filtering by user interests expansion via personalized ranking. IEEE Trans. on Systems, Man and Cybernetics—B, 2012,42(1):218-233.
- [8] Mingdong Tang ; Yechun Jiang ; Jianxun Liu ; Xiaoqing Liu Location-Aware Collaborative Filtering for QoS-Based Service Recommendation . Web Services (ICWS), 2012 IEEE 19th International Conference on Publication Year: 2012 , Page(s): 202 - 209
- [9] Ying-Wei Chen ; Xin Xia ; Yong-Ge Shi. A collaborative filtering recommendation algorithm based on contents' genome. Information Science and Control Engineering 2012 (ICISCE 2012), IET International Conference on . Publication Year: 2012 , Page(s): 1 - 4
- [10] Bobadilla J, Ortega F, Hernando A, Bernal J. Generalization of recommender systems: Collaborative filtering extended to groups of users and restricted to groups of items[J].Expert Systems with Applications. 2012,39 (1):172-86.
- [11] Zhang L, Wu J, Wang Z-C, Wang C-J, editors. Multi-relational Topic Model for Social Recommendation. Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on; 2010: IEEE.
- [12] Bhuiyan T. A survey on the relationship between trust and interest similarity in online social networks[J].Journal of emerging technologies in web intelligence. 2010,2 (4):291-9.
- [13] Ziegler C-N, Golbeck J. Investigating interactions of trust and interest similarity[J].Decision Support Systems. 2007,43 (2):460-75.
- [14] Bhuiyan T. A survey on the relationship between trust and interest similarity in online social networks[J].Journal of emerging technologies in web intelligence. 2010,2 (4):291-9.
- [15] Ziegler C-N, Golbeck J. Investigating interactions of trust and interest similarity[J].Decision Support Systems. 2007,43 (2):460-75.
- [16] Bhuiyan T. A survey on the relationship between trust and interest similarity in online social networks[J].Journal of emerging technologies in web intelligence. 2010,2 (4):291-9.