Journal of Theoretical and Applied Information Technology
31st January 2013. Vol. 47 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN IMPROVED COLLABORATIVE RECOMMENDATION METHOD BASED ON SNA

[1]XU HONG-YAN, [2]LI JUN-PING, [3]FENG YONG
School of Information, Liaoning University, Shenyang Liaoning
E-mail: [1]xuhongyan_lndx@163.com, [2]kayak0123@163.com, [3]fyxuhy@163.com

ABSTRACT

Collaborative recommendation is widely used in e-commerce personalized service, but due to data sparsity and cold start, the existing method cannot give precise results. To improve the recommendation precision, this paper gives a new collaborative recommendation method based on SNA. The proposed method uses social network analysis(SNA) technical to analyze the trust between the users, expresses it as trust value to fill the users—items matrix, and the user similarity calculation has been improved. Finally, an experiment is used to verify the validation of the proposed method. It is proved that it can better solve the problem of data sparsity and cold start.

Keywords: *Social Network; Social Network Analysis (SNA); Similarity; Trust Value; Collaborative Recommendation*

1. INTRODUCTION

Presently, Internet enterprises mainly use collaborative recommendation to develop marketing. Collaborative recommendation uses the history information of users to realize the personalized service. Its core is the collaborative filtering technology [1]. Collaborative recommendation technology mainly includes three types: (1) The collaborative recommendation based on user mainly uses the neighbor user's assessment that has similar interests to generate recommendation to the current user [2]. (2) The collaborative recommendation based on item mainly analyzes the historical information to determine the relationships among different items, and then recommends items by these relationships [3]. (3)The collaborative recommendation based on time-weighted mainly considers the time factor, generally use Ebbinghaus memory curve to measure the influence of time [4]. Internet enterprises often use a technique or a combination of these two techniques according to actual situation, but because all of these techniques have the problem of data sparsity and cold start, so the recommendation accuracy is not high.

Therefore, this paper presents an accurate collaborative recommendation method based on SNA. The method mainly uses SNA technology to analyze the social relationships between users, and then trust degree values are quantized to fill the user-item matrix. This method can solve the data sparsity and cold start, and the accuracy of user's similarity calculation is improved. Experiments show that the method is helpful to improve the precision of collaborative recommendation.

The rest of this paper is organized as follows: SNA and collaborative recommendation were introduced in section 2. The improved method and realization process were introduced in detail in section 3. The contrast experiment was conducted to compare the improved method with the collaborative recommendation based on user and the one based on item in section 4. Results show that the improved method can significantly improve the quality of the recommendation. Finally, the conclusion of this paper is given in section 5.

2. RELEVANT WORK

A. SNA overview

SNA is considered as a structured methodology according to the study on the interaction between social actors [5]. It is mainly to study relationship between the actors on SNA. SNA can analyze the social network from several different perspectives. This paper focuses on the influence of user behavior of the trust relationship among persons. The research shows that there are obvious differences in different types of role, especially in the trust degree. In the study of collaborative recommendation, there is subjectivity, dynamic, unidirectional and other characteristics in users

Journal of Theoretical and Applied Information Technology
31st January 2013. Vol. 47 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

trust. These characteristics play a crucial role in the success of the recommendation, therefore the stranger the trust relationship of the user, the higher the success rate.

B. Collaborative recommendations overview

Collaborative recommendation is to use the similarity between users to generate recommendations, through mastering the relation between individuals to achieve recommendation. The core of the collaborative recommendation is collaborative filtering technology [1]. Its basic idea is that finding groups of users which have the same or similar interests with target users by comparing the user interest and behavior over the degree of similarity, and then predict the target user interest according to their evaluation of resource, which achieves the goal to recommend target users information resources. Collaborative recommendation is divided into the following three steps: (1) establishing a user-relationship matrix, (2) finding the user's nearest neighbor, (3) generating recommendation result set. Finding the user's nearest neighbor is the key of the realization of collaborative recommendation.

3. THE IMPROVED COLLABORATIVE RECOMMENDATION METHOD BASED ON SNA

This paper presents a precise collaborative recommendation method based on SNA. Firstly, trust degree is brought in during the process of forming neighbor set, which not only considers the direct trust relationship of the target users, but also the indirect trust relationship which can be reached by multi steps. Then the similarity calculation is improved based on the user's trust degree which can make the neighbor set constructing more accurate. Finally, the neighbor's score derives predicting score, and get accurate recommended results.

First, we establish a social network graph to use SNA to carry out analysis. Based on users' information we can abstract the relationship between users as a weighted social network chart: $G = (V, E, \lambda)$. Among them, V represents a set of nodes, each node represents all the individual user of the network, E represents the set of edges, it represents the relationship between two individuals (friends, colleagues, family members etc.), λ represents weight value of an edge, which will be defined as the degree of trust in this paper. The network diagram is a directed graph because the trust relation is asymmetric and the edges of the graph have direction.

A. Data Preprocessing

There are two main categories of information recorded by system: one is the information of the user; the other is the information of goods. In this paper, with the help of SNA, recommended method with the trust relationship between users is to be improved. Its basis is that, in real life, people often more easily accept recommendation from the relatives and friends. Therefore, trust and trust degree have important effect on purchase decision-making of target users.

(a) Determination of the user trust relationship

Trust is a kind of faith, which is directed to the ability that an entity can take an action reliably, safely and relying on [5]. It is composed of several attributes, such as integrity, authenticity, reliability, dependability, safety, timely and so on. It chooses the related properties according to the specific environment. The trust relationship of users is mainly determined by the relationship between users; first of all we should make sure that there is relationship between users.

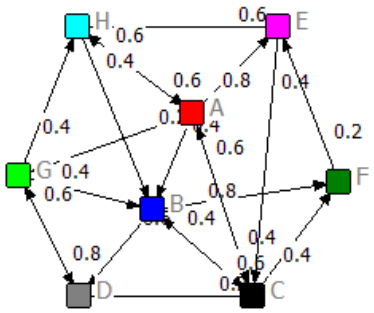

Figure1 User Trust Relationship Diagram

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
A	–	0.4	0.6	0	0.8	0	0.2	0.6
B	0	–	0.4	0.6	0	0.8	0	0
C	0.4	0.6	–	0.2	0	0.4	0	0
D	0	0	0	–	0	0	0.8	0
E	0	0	0.4	0	–	0	0	0.6
F	0	0	0	0	0.2	–	0	0
G	0	0.4	0	0.6	0	0	–	0.4
H	0.6	0.4	0	0	0	0	0	–

Figure2 User Trust Matrix UUA

Between users existence direct relationship and indirect relationship. Direct trust relationship among users is composed of the relevance of users' trust relationship. From a large number of studies of SNA, it is shown that there is the rank in the trust degree of user as follows: the trust of family members and the kinship is the highest, next is friends, classmates, colleagues and neighbors is the lowest. In this paper, we define trust degree in the

Journal of Theoretical and Applied Information Technology
31st January 2013. Vol. 47 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

range of [0, 1] according to five score: the relationships of family and relatives are defined as 1, classmates, friends and colleagues as 0.8; neighbor as 0.6, acquaintances as 0.4; other relationships are from 0.2 to 0. Figure1 is shown the information extracted from some part of the diagram of the user's trust relationships, the node represents the user, the values of the edge defined as user's trust degree .Figure2 is the user trust degree matrix extracted from Figure1.

(b) Relationship between the establishments of the matrix

With constructing the user trust relationship diagram, user trust degree matrix is derived as Figure2 shown. In SNA the transfer of the relationship of users can be measured by matrix multiplication. With $UU_a \times UU_a$ operation, the indirect trust relationship UU' which is reached by two steps can be gotten, $UU_a \times UU_a \times UU_a$ can get indirect trust relationship matrix by tree steps. In this paper we just consider two steps. Each UU'_{ij} of these elements shows that $user_i$ and $user_j$ can reach credibility by two steps. Trust between users should be direct and indirect confidence of credibility, therefore, the ultimate trust degree between $user_i$ to the $user_j$ should be the sum of the trust which should be reached by one or two steps, recorded as *UUA*.

For concerning trust in the same dimension, firstly, standard deviation change of trust matrix is conducted. Calculation method is shown in formula (1):

$$UUA'_{ij} = \frac{UUA_{ij} - \overline{UUA_j}}{S_j} \quad (1)$$

Among them, $\overline{UUA_j} = \frac{1}{n}\sum_{i=1}^{n} UUA_{ij}$, $S_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(UUA_{ij} - \overline{UUA_j})^2}$ represents the transformation of the user's trust . After transformation, each value of the variable is 0, standard deviation is 1. But right now we can't guarantee UUA'_{ij} in the interval [0, 1], so we need carry on the range transformation to get the final trust degree as shown in formula (2):

$$UR_{ij} = \frac{UUA'_{ij} - \min_{1\le i\le m}(UUA'_{ij})}{\max_{1\le i\le m}(UUA'_{ij}) - \min_{1\le i\le m}(UUA'_{ij})} \quad (2)$$

After we got the user trust matrix, consider the amount of target users to its common score project is more than a certain threshold β (β is gotten by the experience) other users of the trust value in the user - item matrix. If the number of projects meet the calculation of the number of neighbors θ (θ is determined by situational) that common score threshold β needs to meet, then we calculate similarity of users with the user trust, get more accurate neighbor set. If the number of projects meeting common score β of neighbors' number is less than θ, then directly combining with user trust matrix. Users, whose target users' trust degree is high, act as its neighbor users directly, and recommend calculation is conducted. This is a very good way to resolve data sparsity and cold start problem, and then gets more accurate recommend results.

B. Users similarity calculation

In the collaborative recommendation, the user similarity calculation is a very crucial step. Similarity calculation method has 3 kinds: cosine similarity, Pearson coefficient similarity and correction cosine similarity [6]. Because cosine similarity did not consider the problem of score scale, the correction cosine similarity is used in this paper. Because we consider the user trust degree can affect the target user's behavior, and we improved calculation formulate based on the trust of users similarity, shown as formula (3).

$$sim(i,j) = U_{ij} \times \frac{\sum_{c\in U_{ij}}(R_{i,c} - \overline{R_i})(R_{j,c} - \overline{R_j})}{\sqrt{\sum_{c\in U_{ij}}(R_{i,c} - \overline{R_i})^2}\sqrt{\sum_{c\in U_{ij}}(R_{j,c} - \overline{R_j})^2}} \quad (3)$$

Among them, the Ri,c and Rj,c respectively represents the ratings of $user_i$ and $user_j$ to mark project c. URij defined as the credibility of $user_j$ that $user_i$ makes. There $\overline{R_i}$ And $\overline{R_j}$ respectively represents the average score of $user_i$ and $user_j$.

C. Predicting commodity score of users

The user similarity is sorted descending, selecting the former n users as the neighbors of target users [7]. Get neighbor set NL, then calculate prediction score for all candidate items by target users, as shown in formula(4):

Journal of Theoretical and Applied Information Technology
31st January 2013. Vol. 47 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$P_{u,c} = \overline{R_u} + \frac{\sum_{i=1}^{n}\left[sim(u,i)\left(R_{i,c} - \overline{R_i}\right)\right]}{\sum_{i=1}^{n} sim(u,i)} \quad (4)$$

Among them, $sim(u,i)$ is the similarity of the $user_u$ and $user_i$ after optimization, $R_{i,c}$ represents the grade that $user_i$ made to $item_c$, $\overline{R_u}$ and $\overline{R_i}$ respectively represent the average score that the $user_u$ and $user_i$ make to the items. Through the formula (4) we could predict the score of the no graded items of target users, the item of the front of the prediction score will be recommend to the target users.

4. EXPERIMENTAL ANALYSES

A. Experimental processing

Background of the experiment is the digital library of school. The digital library in the school system not only includes books information, but also contains user information. The reason to choose digital library as experiment background is the relationship between the users in school environment is simple which can better verify the effectiveness of the proposed method.

The school library users including staff, teachers, students and so on, the relationship between the users can be divided into: colleague relations, teacher relations, students' relations, classmates, friends and other common lending relations, the relationship between strangers. The trust value of student to teacher is 1, colleagues and students is 0.8 friends is 0.6, teacher to students is 0.4, common lending relationship is 0.2, and the other is 0. Then contrast into [0, 1] range. Firstly, trust relationship diagram is set up based on the trust relationship between the users. The point is the individual user, and the value on the edge is the trust value between users.

There are many evaluation standards for recommending quality; this paper mainly uses the mean absolute error (MAE) to recommend quality evaluation. The MAE measures the accuracy of the prediction through calculating the user's prediction score and users' actual deviation between the score, the smaller the MAE recommendation, the higher the quality is. The average absolute deviation MAE calculation is shown as formula (5), among them, the predicted results by the user rating set expressed as $\{p_1, p_2, \cdots p_n\}$, actual user rating for collection is $\{q_1, q_2, \cdots q_n\}$.

$$MAE = \frac{\sum_{i=1}^{n}\left|p_i - q_i\right|}{N} \quad (5)$$

In the experiment, the user's common score amount β is given by experts for 2. θ is a variable, with the amount of θ changing, the proposed method's precision can be further inspected. In the proposed method, the greater the θ is, the more accurate results can be recommended.

B. The results analysis

Compare the method proposed here and the methods from reference [2] and reference [3] in this paper. Use MAE as the recommended quality evaluation criteria, the comparison results are shown in Figure3. We can see from Figure3, with the increase of the number of user neighbors, the given method in this paper has smaller MAE value which is always better than the method in reference [2] and [3]. The method of this paper has a high accuracy recommendation. The reason is that the target user neighbor sets construction is more accurate, and score predicts are more accurately after joining trust relationship between users.

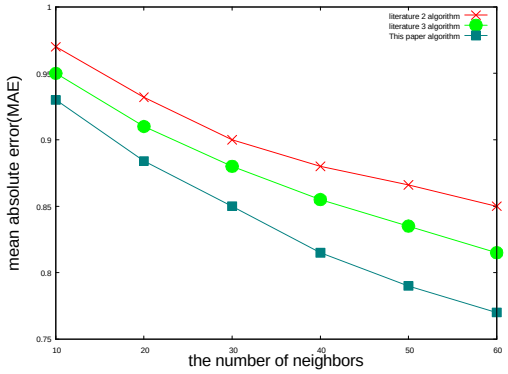

Figure3 MAE Compared Of Recommendation Method

The experimental results show that with the increase of θ, the proposed method greatly improved recommend quality, and solve the sparsity data and cold start problem effectively.

5. CONCLUSIONS

Based on the analysis of trust relationship between users can influence recommendation effectiveness, SNA technology will mix the trust relationship between users into the collaborative recommendation method. A collaborative recommendation method based on SNA is proposed. Experiments show that the proposed method can solve the data sparsity and cold start problem, and considerably improve the quality of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

recommendation, which has high application value. But the premise to use this method is to determine the trust relationship between users, and the complex relationship between users is very hard to gain. This is also the later research emphasis.

6. ACKNOWLEDGMENTS

This work was supported in part by Social Science Youth Foundation of Ministry of Education of China under Grant Nos. 12YJCZH048, the Natural Science Foundation of Liaoning Province of China under Grant Nos. 20102083 and Hundred, Thousand and Ten Thousand Talent Project of Liaoning Province of China.

REFRENCES:

[1] MA HONG WEI, ZHANG GUANG WEI, LI PENG. Collaborative filtering recommendation algorithm review [J]. Journal of Small Miniature Computer Systems, 2009, 30(7): 1282-1288.

[2] LUO XIN, OUYANG YUAN XIN, XIONG ZHANG, et a1. The effect of similarity support in K-Nearest-Neighborhood based collaborative filtering [J]. Journal of Computer, 2010, 8(33): 1339-1445.

[3] DENG AI LIN, ZHU YANG YONG, SHI BO LE. Based on project score predicts the collaborative filtering algorithm recommendation algorithm [J]. Journal of software, 2003, 14(9): 1621-1628.

[4] LIU JIN, WANG QIAN PING, FANG KUN. An optimized collaborative filtering approach combining with item based prediction [C] //Proceeding of the 2007, 11th International Conference on Computer Supported Cooperative Work in Design, 2007: 157-161.

[5] LIU JUN. Introduction to Social Network Analysis [M]. Beijing: Social Sciences Academic Press, 2004: 10-90.

[6] AHN H. J. A New Similarity Measure for Collaborative Filtering to Alleviate the New User Cold Starting Problem [J]. Information Science, 2008, 178(1): 37-51.

[7] SCHAFER J. B., FRANKOWSKI D., HERLOCKER J., SEN S. Collaborative filtering recommender systems [C]. // The Adaptive Web, 2007, Springer LNCS, 4321: 291-324.