http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (22): 7004-7008, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.7004.7008

Using Key Users of Social Networks to Solve Cold Start Problem in Collaborative Recommendation Systems

Zhang Li, Qin Tao and Teng Piqiang
School of Information Technology and Management Engineering,
University of International Business and Economics, Beijing, 100029, China

Abstract: With the application of collaborative filtering technologies and social network in personalized recommendation system, collaborative recommendation techniques based on social network are now made possible. This paper incorporates key users of social network into the traditional collaborative filtering algorithms to solve cold start problem. Also the influence of key users on recommendation accuracy is verified by experiments. Experimental results show that the key users can improve the accuracy of collaborative filtering algorithm which suggests that the key users can be used to alleviate the impact of cold start problem on the recommendation algorithm.

Keywords: Key users, social network, collaborative filtering, cold start

INTRODUCTION

With the development of web2.0 technology and network services, recommendation technologies are applied to various network platforms. Collaborative Filtering (CF) is the most widely used technique in recommendation systems. The main rationale behind CF is that similar users have some interests in common. However, as the number of new users and product items grows rapidly, CF techniques face great challenges in solving cold start problem (Schein *et al*., 2002). On the other hand, studies have shown that, compared to digital or electronic recommendation systems, people prefer personal recommendations from friends (Sinha and Swearingen, 2001).The similarity in background and personal interest helps to build trust between people (Ziegler and Lausen, 2004). The popularity of online social networks motivates researchers to use social relationship attributes to enhance CF algorithm (He and Chu, 2010). So far, relatively little work has been done to fuse the structure characteristics of social network to solve cold start problem in CF. To our knowledge, this paper is among the first to integrate key users of social network into solving the cold start problem in CF.

Literature review: 1 Cold start problem known as new users or new items problem is widely concerned in CF algorithm research. Traditional CF algorithms generally use users' rating information to recommend. As there is no information about new users or items, the traditional CF algorithms are unable to find similar neighbors for new users or items to generate recommendations which is cold-start problem. Research on solving cold start problem can be largely divided into three branches. One stream of research uses a specific method to fill up the new users (items) rating matrix used in the traditional CF algorithm. The mean method, mode method and information entropy method are often used in those studies (Ting *et al*., 2012). However, users' personalized needs are lost in those methods which is contrary to the purpose of personalized information services. Another stream of research first builds a machine-learning model based on user's demographic attributes and item content information to predict the new user or item rating data and then use the traditional CF algorithm to recommend (Cacheda *et al*., 2011; Su and Khoshgoftaar, 2009). For instance, Park and Chu (2009) use regression analysis model to predict the rating. The other stream of research extends the rating matrix using user' attributes and item content information for calculating similarity of users or items. In fact, these methods mainly improve the calculation method of similarity in CF algorithm (Qiu *et al*., 2011; Iaquinta and Semeraro, 2011). Ahn (2008) combined item content information and popularity with user-behavior data to calculate similarity between users. Those methods alleviate the impact of cold-start problem on the CF algorithm performance. But relative to the number of users and items, the dimension of content information or user's attributes is the tip of the iceberg which can't describe user interest effectively. Thus the

Corresponding Author: Zhang Li, School of Information Technology and Management Engineering, University of International Business and Economics, Beijing, 100029, China

recommendation ability of nearest neighbors of the target user obtained by those methods may be greatly reduced. With the development of social network, some researchers improve CF algorithm by integrate friendship, membership or tag data of social network into rating data (Yuan *et al*., 2009; De Meo *et al*., 2011)and solve the cold start problem using the attributes of social network (Sahebi and Cohen, 2011). These studies do not really take advantage of the structural characteristics of the social network. Especially, they do not consider the impact of "excellent" users, who play an essential role in the diffusion of information. Considering all these limitations, in this study, we propose a new algorithm to solve the impact of cold start problem, using information on key users in a social network.

METHOD OF SOLVING COLD START PROBLEM BASED ON KEY USERS

User-based CF is one of most popular method in recommendation systems. We identify it with UCF in this paper. The main steps are as follows: first, calculate the similarity between the target user and other users based on the user-item rating data and select N nearest neighbors of the target user. Then predict the rating of the target user for each item in the common items set rated by the nearest neighbors and recommend the top-M items to the target user based on the predicted rating. The algorithm we propose in this study is mainly improved nearest neighbors selected method, whose basic idea is mining the potential relationship between users based on the extended user-item rating data to build a social network and then identifying key users to select the nearest neighbors for a target user. Other steps of our method are similar to the UCF algorithm.

Social network and key users: A user social network can be built using users' rating data involved in a recommendation system, following this procedure: if user A and B rate the same item, then there is a line (i.e., tie) connecting the two users. The weight (L_{AB}) on the line indicates the degree of association between the two users, as shown in Fig. 1.

We use the Jaccard similarity (Zeng *et al*., 2010) to calculate L_{AB}, shown in Eq. 1, where $I(A)\cap I(B)$ means the common item set rated by user A and B, $I(A)\cup I(B)$ indicates the union item set rated by user A and B, thus $LAB\in[0, 1]$ therefore, a social network composed of nodes and ties is obtained. The greater the weight on the line, the stronger the correlation is between the two users. To reduce the complexity of the network, we only retain the ties whose weights are greater than the threshold value and their associated nodes:

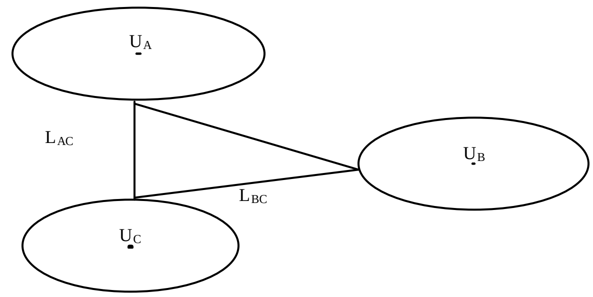

Fig. 1: Social network of users

$$L_{AB}=\frac{|I(A)\cap I(B)|}{|I(A)\cup I(B)|}$$

Key users play an essential role in the diffusion of information. A variety of methods to identify key users have been proposed for social network analysis (Bauer and Lizier, 2012), among which the degree of nodes is the most widely used approach. That is to say key users have higher degree in the social network graph. Also the degree of nodes represents the impact of users (Zeng and Zhang, 2013). In this study, we also use the degree of nodes to identify key users.

Nearest neighbors set: We extend user-item rating data using user's demographic attributes (e.g., age, gender, vocation) and calculate the absolute distance between user i and the new user u as their similarity (Sim(u, i)) based on these attributes. To verify the effect of key users on the performance of user-based CF algorithm, we construct the nearest neighbors set N_u for the target user u in two ways:

One way counts nodes (i.e., users) degree in the social network (shown in Fig. 1) and selects top-N1 users with highest degree to build a candidate nearest neighbors set. And then calculates the similarity between the new user and users in the candidate set. At last N nearest neighbors are selected to construct N_u for recommendation.

The other way calculates the similarity between the new user and other users and selects N1 nearest neighbors to build a candidate set. And then counts the nodes (i.e., users) degree in the candidate set based on the social network (shown in Fig.1). At last selects top-N users with highest degree to construct nearest neighbors set Nu for recommendation.

The above two ways are mainly about nearest neighbors selected method which integrate key users into the nearest neighbors set. And their biggest difference is the using timing of key users. Once the nearest neighbors set for the target user established by the above two ways, we can use it to recommend. The other steps of our recommendation process are similar to the UCF algorithm. So we identify the CF algorithm based on the above two ways with KUCF1 and KUCF2 respectively.

Predicted rating: We use the weighted average method to predict the rating of a new user [2], as shown in Eq. 2:

$$p_{u,t} = \overline{R_u} + \frac{\sum_{i\in N_u} \text{sim}(u, i)\times(R_{i,t} - \overline{R_t})}{\sum_{i\in N_u} \text{sim}(u, i)} \quad (2)$$

where N_u is the nearest neighbor set of new user u, $p_{u,t}$, is the predicted rating of user u for item $t(t\in I(N_u)$ which is the common item set rated by user in N_u) and $\overline{R_t}$ is the average rating of user i.

EXPERIMENTS

The MovieLens dataset is used in our experiments which is a popular dataset used by researchers and developers in the field of collaborative filtering. The dataset (100 KB) contains data on movie ratings provided on the MovieLens website of movie recommenders. From these data, users with less than 20 ratings were removed, resulting in a total of 100,000 ratings from 943 users on 1,682 movies. We extend user-item rating data using user's demographic attributes (e.g. age, gender, vocation). In our experiments, all data were standardized into [0, 1], shown in Table 1. We select 100 users in 943 users as new users.

Evaluation metrics: In this study, we use the recommendation accuracy as the measure for performance evaluation which considers the consistency of two top-M items list recommended by the algorithm and rated actually by the target user. In our method, we don't consider the items order in the lists. So for the target user i, calculate its recommendation accuracy $P_i(M)$ by Eq. 3:

$$P_i(M) = \frac{R_i(M)}{M} \quad (3)$$

where, R_i (M) represents the number of item in the intersection of two lists and M is the number of recommended item. The average recommendation accuracy of all target users P (M) can be calculated by Eq. 4:

$$P(M) = \frac{i}{m}\sum_{i=1}^{m} P_i(M) \quad (4)$$

where, m is the number of target user set m = 100 in the study.

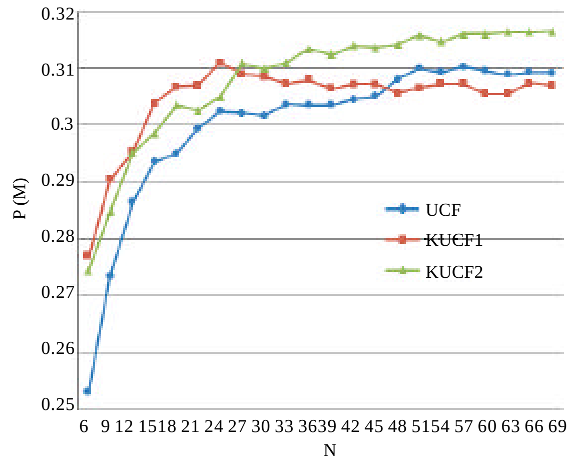

Fig. 2: P(M)Comparison of the UCF, KUCF1 and KUCF2 Algorithm

Table 1: User-item extended rating matrix

	Item							
User	$Item_1$	$Item_2$	$Item_3$	…	$Item_n$	Age	Gender	Vocation
$User_1$	0.6		0.8	…	0.8	0.4	1.0	0.7
$User_2$	0.8	0.6		…	1.0	0.3	0.0	0.2
…	0.8		0.8	…	1.0	0.7	0.0	0.9
$User_n$	1.0	1.0	0.4	…		0.3	1.0	0.1
New user			…		0.4	1.0	0.4	

Experimental results: To test the performance of the method proposed, we do three experiments using KUCF1, KUCF2 and UCF algorithm respectively based on the above data. The results of P(M) (M = 110) are shown in Fig. 2, where N is the number of nearest neighbors. Also the experimental results were standardized into [0, 1].

From Fig. 2, we can see the accuracy of KUCF1 is better than that of UCF while N[6, 45]. But when N is greater than 45, the accuracy of UCF algorithm is better and the two curves become stabilized. According to the basic rationale of collaborative filtering, the more similarity between the target user and its neighbor, the greater value the neighbor holds for recommendations. So we analysis the average similarity between the target users and users in the nearest neighbors set used in UCF and KUCF1 algorithm by formula (5), to find the reasons for the two curves intersect in Fig. 2:

$$avg_{sim}(N) = \frac{\sum_{i=1}^{N}\sum_{u=1}^{m}\frac{\text{sim}(u, i)}{N}}{m} \quad (5)$$

where, m represents the number of target user and N represents the number of the nearest neighbor.

The experimental results are shown in Fig. 3. As can be seen from Fig. 3, the average similarity between the

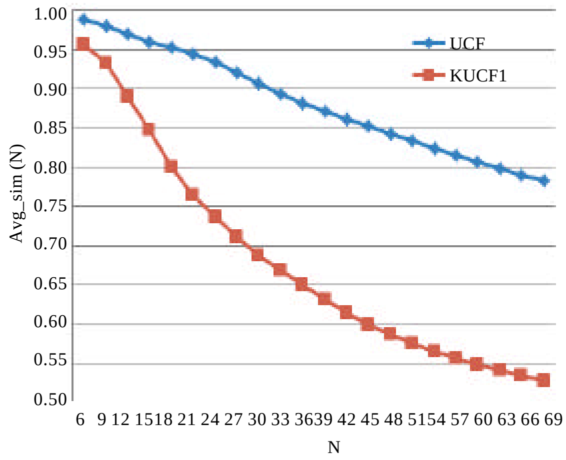

Fig.3: The average similarity in UCF and KUCF1 algorithm

target users and users in the nearest neighbors set used in KUCF1 is less than what in UCF and as N increases, the gap is expanded. According to the basic rationale of collaborative filtering, we think that is resulting in KUCF1 recommended effect reduced. This also verifies the similarity of users affects the performance of CF algorithm more.

Yet the accuracy of KUCF2 is always superior to that of UCF in Fig. 2 which also verifies that integrate key users into the nearest neighbors set used in CF can enhance the accuracy of algorithm. However, identifying key users increases the complexity of the algorithm, making it more time-consuming to generate recommendations. We may identify key users offline, so it will not compromise the promptness of making recommendation.

CONCLUSION

The popularity of online social network motivates researchers to improve the traditional CF algorithms using the social attributes of a network. Key users play an indispensable role in the dissemination of information. This study integrates key users of a social network into the CF algorithm to solve cold start problem. To a certain extent, the method proposed in this study testifies the impact of key users on the performance of the CF algorithm. Further research may take into account the problem of users' interest shifting which we didn't address in this study.

ACKNOWLWDGMENTS

This study was supported by National Social Science Foundation of China (Grant No. 13BTQ027), Beijing Philosophy and Social Science Planning Foundation (Grant No. 12JGB034). We are grateful to the editors and the reviewers whose comments have improved this paper considerably.

REFERENCES

Ahn, H.J., 2008. A new similarity measure for collaborative filtering to alleviate the new user Cold-starting problem. Inform. Sci., 178: 37-51.

Bauer, F. and J.T. Lizier, 2012. Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: A walk counting approach. Europhysics Lett., Vol. 99. 10.1209/0295-5075/99/68007

Cacheda, F., V. Carneiro, D. Fernandez and V. Formoso, 2011. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. ACM Trans. Web, 5: 2-33.

De Meo, P., E. Ferrara, G. Fiumara and A. Provetti, 2011. Improving recommendation quality by merging collaborative filtering and social relationships. Proceedings of the 11th International Conference on Intelligent Systems Design and Applications, November 22-24, 2011, Cordoba, pp: 587-592.

He, J. and W.W. Chu, 2010. A Social Network-Based Recommender System. Springer, US., pp: 47-74.

Iaquinta, L. and G. Semeraro, 2011. Lightweight approach to the cold start problem in the video lecture recommendation. Proceedings of the ECML/PKDD Discovery Challenge Workshop, Volume 770, September 5, 2011, Athens, Greece, pp: 83-94.

Park, S.T. and W. Chu, 2009. Pairwise preference regression for cold-start recommendation. Proceedings of the 3rd ACM Conference on Recommender Systems, October 23-25, 2009, New York, pp: 21-28.

Qiu, T., G. Chen, Z.K. Zhang and T. Zhou, 2011. An item-oriented recommendation algorithm on cold-start problem. EPL, Vol. 95. 10.1209/0295-5075/95/58003

Sahebi, S. and W.W. Cohen, 2011. Community-based recommendations: a solution to the cold start problem. Proceedings of the Workshop on Recommender Systems and the Social Web, October 23-27, 2011, Chicago, IL., USA.

Schein, A.I., A. Popescul, L.H. Ungar and D.M. Pennock, 2002. Methods and metrics for Cold-start recommendations. Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 11-15, 2002, Finland, pp: 253-260.

Sinha, R. and K. Swearingen, 2001. Comparing recommendations made by online systems and friends. Proceedings of the DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries, Volume 106, June 18-20, 2001, Dublin, Ireland.

Su, X. and T.M. Khoshgoftaar, 2009. A survey of collaborative filtering techniques. Adv. Artificial Intell., 10.1155/2009/421425

Ting, S.D., H. Tao and Z.F. Hai, 2012. Survey of Cold-start problem in collaborative filtering recommender system. Comput. Modernization, 5: 59-63.

Yuan, Q., S. Zhao, L. Chen, Y. Liu, S. Ding, X. Zhang and W. Zheng, 2009. Augmenting collaborative recommender by fusing explicit social relationships. Proceedings of the Workshop on Recommender Systems and the Social Web, Octobet 25, 2009, New York, pp: 49-56.

Zeng, A. and C.J. Zhang, 2013. Ranking spreaders by decomposing complex networks. Phys. Lett. A, 377: 1031-1035.

Zeng, W., M.S. Shang, Q.M. Zhang, L. Lue and T. Zhou, 2010. Can dissimilar users contribute to accuracy and diversity of personalized recommendation? Int. J. Modern Phys. C, 21: 1217-1227.

Ziegler, C.N. and G. Lausen, 2004. Analyzing Correlation between Trust and User Similarity in Online Communities. Springer, Berlin Heidelberg, pp: 251-265.