AMSE JOURNALS-2016-Series: Advances B; Vol. 59; N°1; pp 62-75

Submitted July 2016; Revised Oct. 15, 2016, Accepted Dec. 10, 2016

Research on Personalized Tourism Attractions Recommendation Model Based on User Social Influence

*Zhijun Zhang, **Huali Pan, *Gongwen Xu

* School of Computer Science and Technology, Shandong Jianzhu University

China, Jinan 250101, (zzjsdcn@163.com)

** Business School, Shandong Normal University

China, Jinan 250014, (panhuali@sina.com)

* School of Computer Science and Technology, Shandong Jianzhu University

China, Jinan 250101, (xugongwen@sdjzu.edu.cn)

Abstract

With the rapid development of social networks, location-based social network gradually rise. In order to retrieve user most prefer attractions from a large number of tourism information, location-based personalized recommendation technology has been widely concerned in academic and industry. For the solving techniques problems such as data sparsity and cold-start existed in personalized recommendation system, this paper proposes a personalized location recommendation model-SocInflu, which combines user collaborative filtering technology with social networks factor. This method fully exploits social relations and trust relations between users and recommends users most interest attractions by the means of social influence and position information between user and tourism attractions. Experimental results on real data sets demonstrate the feasibility and effectiveness of the proposed model. Compared with the existing recommendation algorithm, it has higher prediction accuracy.

Key words

Social influence, Personalized recommendation, Collaborative filtering, Social network, Preference prediction

1. Introduction

With the rapid development of Internet, the structure of travel website becomes more and more complex. At the same time, it provides users with more and more tourism information. How to find the exact information immediately from vast amounts of network information for tourist has become more and more difficult [1]. Recommended systems can provide travelers with filtered information by predicting attractions preference degree of travelers and apply knowledge discovery technology to generate personalized attractions recommendation to help travelers find the exact information what they need. With the development and popularization of personalized service, recommendation system is widely used in various tourism web sites. With good development and application prospects, recommendation systems have become an important research direction in the web intelligence technology. It has been widely concerned by many researchers in academic and industry.

Personalized recommendation is applied to the field of tourism, including recommended both individual and integrated tourism products. The basic principle of recommendation system is: firstly, save users' historical behavior data, such as browsing, buying, comments and rating, etc.; secondly, dig user preference information based on these records and then analyze user preferences and construct user preference model. Collaborative filtering (CF) recommendation system could automatically predict the current user's preferences by collecting some evaluation information from other users or similar items. Collaborative filtering has been widely used in some large and well-known business systems, such as Amazon and Alibaba [2]. Currently, the collaborative filtering algorithm includes memory-based, model-based and mixed two recommended techniques [3]. The nature of traditional collaborative filtering algorithms is using the interaction between individuals within the group to make predictions for the properties of current object, but the prediction is unidirectional, because the prediction only consider the influence of other individuals on the current object, and without taking into account the current object also has a certain influence on other individuals in the group [4]. This paper is inspired from the concept of influence sets in the field of information retrieval. Traditional collaborative filtering algorithm is incomplete. It is necessary to find the group affected by the current target groups, combined with the properties of the two groups together to make a determination of the properties of the current object.

In the social networks, user's interests and preferences often influenced by friends or friends of friends, which is the so-called social influence. Some of the existing recommendation algorithm are based on user interest preferences or based on the similarity between items, but these algorithms are not really to expose attribute information about user and the project itself.

They don't deeply analyze users' social influence, and therefore they can not accurately recommend users' interested in products. To solve this problem, this paper proposes a personalized tourism attraction recommendation algorithm-SocInflu, which is a combination algorithm of collaborative filtering and social influence factor.

2. Related work

As location-based social networking (LBSN) being developed rapidly, social web sites such as Foursquare, Gowalla, and Flickr emerged. Travelers like to upload their tourism attractions and photos to social web sites to share their travel experiences, or release check in information to share about their location and service. In order to minimize the risk of the transaction, travelers often inquire their friends or colleagues about the integrity degree or other aspects of the scenic spot. In real life, the reputation of a landscape often is social networks that are made up of the relationship of blood, friends, and colleagues between travel enthusiast, which is evaluated and passed from each other as a form of word-of-mouth. The trust to a scenic spot can be passed through social networks to other travel enthusiasts, so a trust network with the function of recommendation is formed.

In order to find the role of social influence on personalized recommendation, scholars carried out extensive research and exploration on the formation group and behavior preferences in social network from aspects of both theoretical and practical. In 1994, Wasserman [5] analyzed social network and found that not only relations could be formed between the two users, and these relations might form clusters relationship. Earlier studies about social influence on human behavior and social life are the Project of Digital Youth [6], this project made a quantitative analysis and research on the power of American teens joining online social networks (OSN). Backstrom[7] studied the individuals' behavior of joining, growth and evolution in the community in social network, and found an individual joining the community, which is usually influenced by his friend. Anagnostopoulos[8] studied the influence and relationships in social networks, defined several general models, and further exposed the generation of social impact and its impact on other members. Marlow [9] proposed to use label study the relationship between users and their friends in Flickr and found the relationship between social relations and words labels. Ma [10] studied how to use the trust relationship between users to further improve the performance of traditional recommendation algorithm, and gave a probability matrix factorization framework that could merge trust relationship information [11]. Noor [12] proposed personalized recommendation system that combine social networking technologies and semantic web, which built a relation between user preferences and personalized search. Shi [13] studied the behavior pattern of user, found that users' behavior characteristics have a strong law in the different data sets and online collaborative filtering method could compound new data. Khoshneshin [14] proposed an improved co-clustering method, which maintained the stability. This method greatly improved the accuracy of prediction by joint parallel clustering operation to solve the time-sensitive issues. Zhang [15] proposed a personalized recommendation algorithm that combined the temporal behavior and trust relations, and could improve user satisfaction. Works of these scholars are trying to solve the contact problem between user preferences and recommended item in personalized recommendation.

For some of the weaknesses of the traditional user modeling techniques, this paper makes an exploratory research for user's social influence in personalized recommendation system, proposes personalized recommendation model and algorithms based on social influence. The results in real data sets show the proposed SocInflu algorithm in this paper is superior to the traditional collaborative filtering algorithms either in running efficiency or in recommendation accuracy.

3. Recommendation model based on users' social influence

3.1 Diagram of social network

In recent years, with the rise of such Faeebook, Twitter and other social media, there commendation methods that using social relationships among users have gradually become a research hot spot in recommendation field [16] Such methods assume that the user decision-making process is vulnerable to trust relationship or friendship in the recommendation process, and friends trusting each other have similar interests. This paper gives a general definition of trust-based social network diagram.

Assuming directed social network diagram $G=(U \cup V, E)$, the set of vertices $U=\{u_i\}_{i=1}^{N}$ represents all users in the social network, the set of vertices $V=\{\mathrm{v}_j\}_{j=1}^{M}$ represents all items in social network, E is set of edges, which represents all relationships of user or item. Assuming $u_1,u_2 \in U$ represents user, $\mathrm{v}_1,v_2 \in V$ represents item; edge $(u_1,u_2) \in E$ represents user u_1 trusts user u_2, edge $(v_1,v_2) \in E$ expresses items v_1 and v_2 being bought by the same user in a certain period of time; $u \in U$ represents user, $\mathrm{v} \in V$ items; edge $(u,v) \in E$ represents user u rating to item v; Symbol N_u indicates neighbor set of user u, symbol N_{i} indicates neighbor set of item i.

Anagnostopoulos proposed concepts and definitions of social network diagram based on influence[17]. Assuming α to be individuals in a social network, the definition of probability that α is influenced by its activated friends is shown in Equation (1).

$$p(a) = \frac{e^{\alpha \ln(\alpha+1)+\beta}}{1+e^{\alpha \ln(\alpha+1)+\beta}} \tag{1}$$

In Equation (1), α and β is correlation coefficient. Equation (1) can be written as equivalent to the following Equation (2).

$$\ln(\frac{p(\alpha)}{1-p(\alpha)}) = \alpha \ln(\alpha+1)+\beta \tag{2}$$

Coefficient α is used to measure social relevance, the greater of its value, represents more relevant. The value of α and β is estimated by maximum likelihood logistic regression function. $Y_{a,t}$ represents the number of user that there are a activating friend and activated at initial time ; $N_{a,t}$ represents the number of user that there are a activating friend and inactivated at that moment.

$$Y_a = \sum_t Y_{a,t} \tag{3}$$

$$N_a = \sum_t N_{a,t} \tag{4}$$

The values of α and β can be obtained by maximized Equation (5) .

$$\prod_a p(a)^{Y_a}(1-p(a))^{N_a} \tag{5}$$

3.2 User-location check-in frequency matrix

The situation of check-in between user-location can be expressed by user-location check-in frequency matrix [18] as shown in table 1.

Table 1: User-location check-in frequency matrix

User/Loction	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**
U1	?	?	10	?	?	2
U2	1	?	?	25	?	7
U3	?	3	35	?	?	1
U4	2	?	?	9	?	?
U5	?	54	?	?	68	3

3.3 Social influence factor

Document [19] proposed that social influence factor between friends is also associated with the number of their mutual friend. Social influence factor among users can be calculated by whether they are friend and the number of their mutual friend, the computational formula is shown in Equation (6).

$$SI_{k,i} = \eta \cdot f_{k,i} + (1-\eta) \cdot \frac{|F_k \cap F_i|}{|F_k \cup F_i|} \tag{6}$$

In Equation (6), η is an adjustable parameter, its range is [0,1]. $f_{k,i}$ indicates that whether there is a friend relationship between user u_k and u_i. $f_{k,i}=1$ represents users u_k and u_i are friends, $f_{k,i}=0$ represents users u_k and u_i are not friends. F_k represents friends of user u_k dataset. Parameter η needs adjust in this algorithm. Adjusting parameter helps to make the algorithm t better. η balanced the weight of friends and potential friends, if η is bigger, recommendation is better, represents that friends have a greater impact on user check-in habits user than potential friends; if η is bigger, recommendation is worse, represents that friends have a less impact on user check-in habits user than potential friends.

When $\eta=1$, Equation (1) can be transformed into Equation (7).

$$SI_{k,i} = f_{k,i} \tag{7}$$

When $\eta=0$, Equation (1) can be transformed into Equation (8).

$$SI_{k,i} = \frac{|F_k \cap F_i|}{|F_k \cup F_i|} \tag{8}$$

3.4 Calculating Model for user interest degree to candidate location-SocInflu

Whether in traditional social network or in location-based social network, friends often have similar behavior. Because they are friends, so they may have a lot of common interests, and have related check-in behavior. For example, we will travel with friends and share current status together. Or we see friends share good restaurants, so we are likely to go there to eat next time. This shows that using social network influence can improve the quality of places recommendation [20] Specifically, if user wants to go to a strange place, his friend is familiar with that place, so his friend's recommendation is more valuable to him.

After using Equation (6) to calculate the similarity $SI_{k,i}$ combined social influence factor and user check-in, select $top-N$ of $SI_{k,i}$, users with highest rating make up user sets $U^{'}$ and $U^{'} \in U$, $U^{'}$ expresses user set of $top-N$ user similar with user u_i. For the place users have not go, to calculate user's interest degree, specific calculating formula is shown in Equation (9).

$$p(u_i, l_j) = \frac{\sum_{u_k \in U^{'}} SI_{k,i} \cdot c_{k,j}}{\sum_{u_k \in U^{'}} SI_{k,i}} \tag{9}$$

The SocInflu method does not increase the time complexity much compared with every single recommendation method since the rating matrix are very sparse.

3.5 System structure figure

The system includes two parts: server and client, as shown in Figure 1. Server stores user check in data and analyzes the data, running travel package recommendation algorithm, in response to a request sent by client. Client provides user interface, the user submits the travel demand through the client, browses the recommendation result.

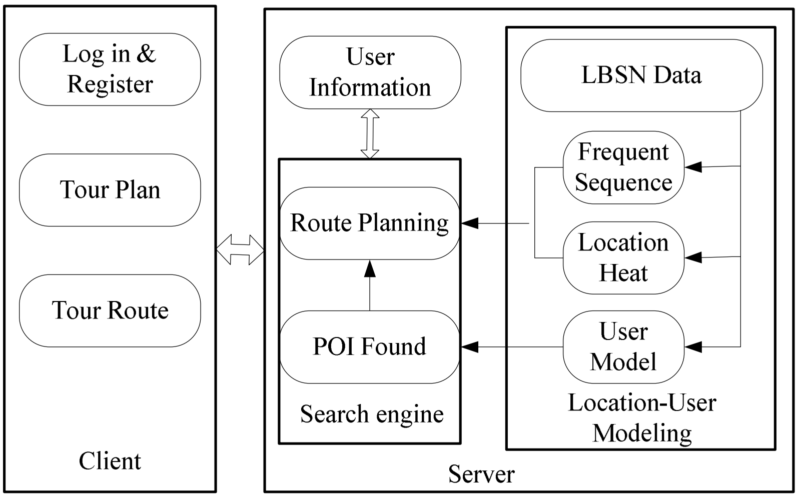

Figure1: System structure figure

4. Experimental results and analysis

4.1 Introduction of Dataset

The experiment uses Foursquare data sets. Foursquare data sets include user's check-in records, friends' relations, habitual residence. Foursquare website provides user-targeted social networking services, encourages mobile phone users to share their current geographical location.

In which, check-in data has more than 100 million, the number of users is more than 10,000 in friendships data, and friendships is more than 40,000.

4.2 Comparative method

To test the performance of proposed SocInflu recommendation model in this paper, we do experiments to verify the validity of the model. We select collaborative filtering algorithm UserCF [21] as a reference model. UserCF algorithm mainly recommends tourist attractions or commodities by finding similar users. The similarity between users is calculated by Pearson similarity Equation (10).

$$sim(u,a) = \frac{\sum_{i \in I_{u,a}} (R_{u,i} - \overline{R}_u)(R_{a,i} - \overline{R}_a)}{\sqrt{\sum_{i \in I_{u,a}} (R_{u,i} - \overline{R}_u)^2} \sqrt{\sum_{i \in I_{u,a}} (R_{a,i} - \overline{R}_a)^2}} \tag{10}$$

Where, $sim(u,a)$ is the similarity of user a and u. $I_{u,a}$ represents product group rated by u and a, $R_{u,i}$ and $R_{a,i}$ represents rating of item i from u and a, $\overline{R}_u$ and $\overline{R}_a$ represents respectively the mean value of all goods from u and a.

4.3 Metrics

When websites providing recommendation services, typically give users a personalized recommendation list, this is called Top-N recommendation. Prediction accuracy of Top-N recommendation is generally measured by Precision and Recall.

R (u) is the recommendation list based on user behavior on the training set, and T (u) is a list of user behavior on the test set. Recall of recommendation result is defined as Equation (11).

$$Recall = \frac{\sum_{u \in U} | R(u) \cap T(u) |}{\sum_{u \in U} | T(u) |} \tag{11}$$

Precision of recommended result is defined as Equation (12).

$$Precision = \frac{\sum_{u \in U} | R(u) \cap T(u) |}{\sum_{u \in U} | R(u) |} \tag{12}$$

In order to compare the performance of proposed algorithm, this paper uses Precision to evaluate the performance of recommendation algorithm. The value of Precision is greater, the predict is more useful.

4.4 Experimental results

4.4.1 Relationship between check-in location and distance of habitual residence

This paper studies the effects of distance of habitual residence and location on check-in possibilities in Foursquare data[22], as shown in Fig 2. Where the horizontal axis represents the distance from check-in location to habitual residence, the vertical axis represents check-in record proportion at this distance.

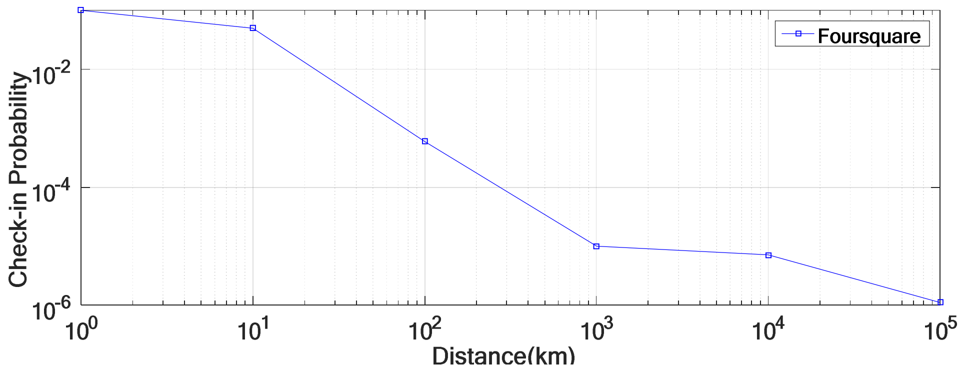

Fig. 2 Effects of distance on check-in possibilities

Seen from Figure 2, distribution function shows as power-law function. When the distance is less than 1000 km the distribution decreases rapidly. When the distance is more than 1000 km, the distribution decreases slowly. The distance between user's habitual residence and his check-in location is generally short, which makes location class cluster phenomenon nearby habitual residence. So you can find the user's habitual residence by check his check-in record. In general, the place with most check-ins can be inferred for the user's habitual residence. Meanwhile, at the place far away, there is a possibility of high check-ins, which is the result of a user traveling to different places.

4.4.2 The impact of social influence on accuracy of recommendation

The idea of verifying algorithm: remove some of the check-in locations in the data set, and make these data as a monitoring point, then using the remaining data to verify each recommendation algorithm. At last to calculate accuracy of the recommended algorithm to the user by the monitoring point appears at the final recommended location.

Due to limited by human activity patterns, we need to filter out the place that is far away from the user's current location. For example, a user travel in Shanghai, if the system recommends some attractions in Qingdao to him, he is certainly not to accept the recommendation. Therefore, we propose a distance variable D. If the distance between recommended location and the user's current position is greater than D, then filter out the place. We can set the variable D according to application requirements. In this paper, we do experiments by setting D as 5km and 20km respectively. Firstly, we do experiment by setting the distance variable D as 5km. When calculating Precision @ 5, the effect of the entire recommendation algorithm is shown in Fig 3.

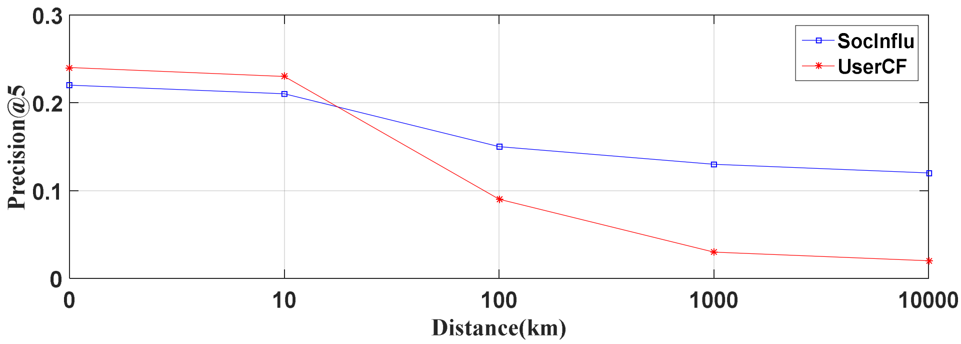

Fig.3 Comparison of recommendation algorithm Accuracy-Precision@5

As it can be seen from Figure 3, when the distance exceeds 20km, proposed SocInflu algorithm in this paper always has the best recommendation effect, which indicating that these two factors-user preferences and social impact can improve the recommendation algorithm accuracy. Then, we do experiment by setting the distance variable D as 10km. When calculating Precision@10, the effect of the entire recommendation algorithm is shown in Fig 4.

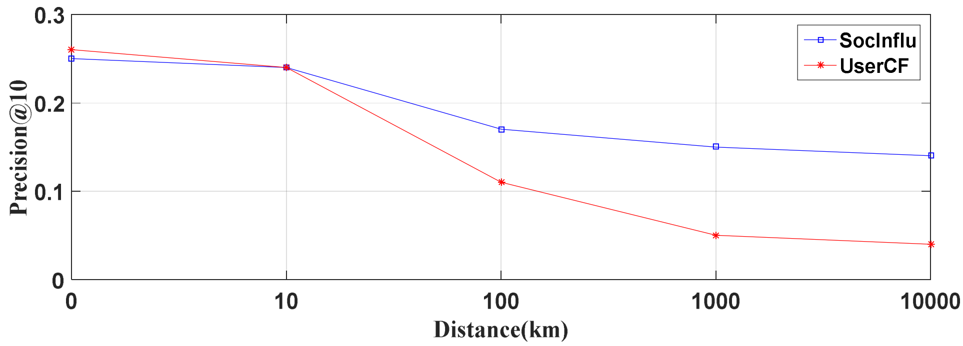

Fig. 4 Comparison of recommendation algorithm accuracy-Precision@10

As can be seen from the comparison of Figures 4 and 3, almost all of the recommendation algorithm accuracy of Precision@10 is higher than that of Precision @5, the reason is when the base number is larger, the recommendation algorithm accuracy is higher.

4.4.3 The Impact of Distance Variable D on Recommendation Accuracy

Next, we discuss when set the distance variable D as different value, how it effects on recommendation algorithm accuracy. On SocInflu algorithm for Precision@5, we discuss the impact of distance variable D on recommendation accuracy. The experiment results are shown in Fig 5.

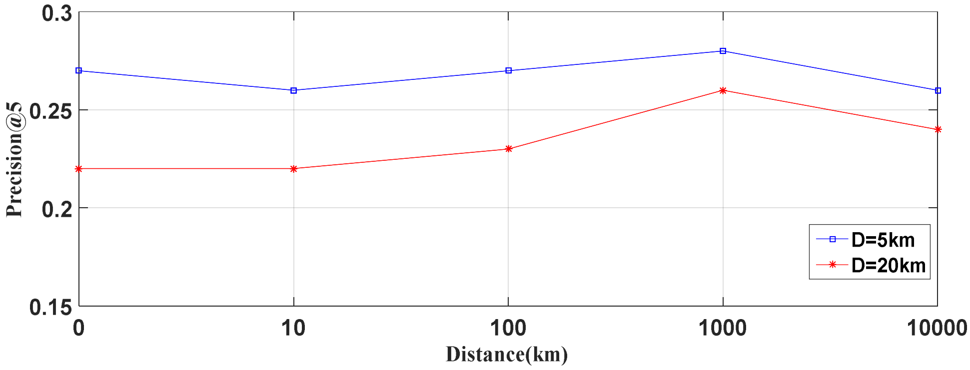

Fig. 5 The impact of distance variable D on recommendation accuracy

As it can be seen from Figure 5, when $D = 20km$, the efficiency of the SocInflu algorithm is better than that of $D = 5km$. Also we can see that when the distance from habitual residence is more than 1000 kilometers, the effect of D on the accuracy is little.

5. Discussion

This paper makes an exploratory research for user's social influence in personalized recommendation system, proposes personalized recommendation model and algorithms based on social influence. The contribution of this study can be explained as follows. We deeply analyze users' social influence, and therefore the proposed method can accurately recommend users' interested in POI. Our findings show that trust and social influence also have positive relationships with reuse intentions with regard to a tour web service provider. Previous studies have suggested that the usefulness and accuracy of recommender systems have positive effects on reuse intentions. We added considerations of the effects of social influence on evaluating recommender systems that involve other users.

Nonetheless, this study showed the importance of social influence in personalized recommender systems. Managers of online stores providing personalized recommendations should focus on the experience of social influence in interactions with recommendation systems to enhance trust placed in such systems. Although some Web 2.0 applications provide information about similar users for specific items, many transactional online stores continue to provide recommendations without generating any experience of social influence. Many social networking sites seek to exploit social network information for purposes of services or products with greater appeal. Social influence on Web services may enhance the effectiveness of their activities.

There are several directions worthy of considering for future study: 1) how to model extremely sparse frequency data, e.g., by designing more subtle sampling techniques, to improve my methods; 2) how to include other information, e.g., location category, and activity, into our fused framework; 3) how to incorporate temporal effect on POI recommendation to capture the change of users' preference.

6. Conclusion

This paper introduced position social networks and associated recommendation algorithm, studied the influence of user preferences and social factors on location recommendation. In addition, the paper analyzed the problems faced by the personalized location recommendation algorithm, and proposed a personalized location recommendation model-SocInflu that combines collaborative filtering with social networks. Finally, doing experiments shows the feasibility of this idea, and the experimental results show that the improved algorithm can indeed solve the problems faced by the personalized location recommendation. However, this paper only studied attractions recommendation. For friend recommendation and activities recommendation, this paper just introduced the current application researches, which are important part of social recommendation system and are also worthy of further study.

Acknowledgments

This paper is supported by the Science and Technology Development Planning of Shandong Province (2014GGX101011, 2015GGX101018), A Project of Shandong Province Higher Educational Science and Technology Program (J12LN31, J13LN11,and J14LN14) and Jinan Higher Educational Innovation Plan (201401214, 201303001).

References

1. Shen P, Wang H, Meng Z, et al. An Improved Parallel Bayesian Text Classification Algorithm[J]. Review of Computer Engineering Studies, 2016, 3(1): 6-10.

2. Zhang Z, Liu H. Application and Research of Improved Probability Matrix Factorization Techniques in Collaborative Filtering[J]. Int J Control Autom, 2014, 7(8):79-92.
3. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. Knowledge and Data Engineering, IEEE Transactions on, 2005, 17(6): 734-749.
4. Chen J, Yin J. A collaborative filtering recommendation algorithm based on influence sets. Journal of Software, 2007, 18(7):1685−1694.
5. Wasserman S, Faust K. Social network analysis: Methods and applications[M]. Cambridge university press, 1994.
6. Boyd D M. Taken out of context: American teen sociality in networked publics[M]. ProQuest, 2008.
7. Backstrom L, Huttenlocher D, Kleinberg J, et al. Group formation in large social networks: membership, growth, and evolution[C]//Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006: 44-54.
8. Anagnostopoulos A, Kumar R, Mahdian M. Influence and correlation in social networks[C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008: 7-15.
9. Marlow C, Naaman M, Boyd D, et al. HT06, tagging paper, taxonomy, Flickr, academic article, to read[C]//Proceedings of the seventeenth conference on Hypertext and hypermedia. ACM, 2006: 31-40.
10. Ma H, King I, Lyu M R. Learning to recommend with social trust ensemble[C]//Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009: 203-210.
11. Ma H, Yang H, Lyu M R, et al. Sorec: social recommendation using probabilistic matrix factorization[C]//Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008: 931-940.
12. Noor S, Martinez K. Using social data as context for making recommendations: an ontology based approach[C]//Proceedings of the 1st Workshop on Context, Information and Ontologies. ACM, 2009: 7.
13. Shi X, Zhu J, Cai R, et al. User grouping behavior in online forums[C]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009: 777-786.

14. Khoshneshin M, Street W N. Incremental collaborative filtering via evolutionary co-clustering[C]//Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010: 325-328.
15. Zhang Z, Liu H. Social recommendation model combining trust propagation and sequential behaviors[J]. Applied Intelligence, 2015, 43(3): 695-706.
16. Guo L, Ma J, Chen ZM. Trust Strength Aware Social Recommendation Method [J]. Journal of Computer Research and Development, 2015, 50(9): 1805-1813.
17. Anagnostopoulos A, Kumar R, Mahdian M. Influence and correlation in social networks[C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008: 7-15.
18. Cheng C, Yang H, King I, et al. Fused matrix factorization with geographical and social influence in location-based social networks[C]//Twenty-Sixth AAAI Conference on Artificial Intelligence. 2012.
19. Konstas I, Stathopoulos V, Jose J M. On social networks and collaborative recommendation[C]//Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009: 195-202.
20. Cho E, Myers S A, Leskovec J. Friendship and mobility: user movement in location-based social networks[C]//Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011: 1082-1090.
21. Li Y,Chen C,Qi L. Collaborative Filtering based on User Behavior Similarity [EB/OL]. http://www.paper.edu.cn/releasepaper/content/201009-117.
22. Zhang ZJ, Liu H. Research on context-awareness mobile SNS recommendation algorithm[J]. Pattern Recognition and Artificial Intelligence, 2015, 28(5): 404-410.