An Integrated Recommendation Approach Based on Influence and Trust in Social Networks

Weimin Li[1,2], Zhengbo Ye[1,2], and Qun Jin[3]

[1] School of Computer Engineering and Technology, Shanghai University, Shanghai, China
[2] Shanghai Key Laboratory of Computer Software Evaluating and Testing, Shanghai, China
[3] Graduate School of Human Sciences, Waseda University, Tokorozawa, Japan
{wmli,yezb}@shu.edu.cn, jin@waseda.jp

Abstract. In real human society, influence on each other is an important factor in a variety of social activities. It is obviously important for recommendation. However, the influence factor is rarely taken into account in traditional recommendation algorithms. In this study, we propose an integrated approach for recommendation by analyzing and mining social data and introducing a set of new measures for user influence and social trust. Our experimental results show that our proposed approach outperforms traditional recommendation in terms of accuracy and stability.

Keywords: recommendation algorithms, similarity, influence, social trust.

1 Introduction

In recent years, recommender systems have been an important tool to help people to get information that meets their needs or interests from the mass data. Traditional recommendation algorithms generally make use of the users' behavior data or attribute data. As one of the most successful recommendation algorithms in commercial domains, collaborative filtering is designed through computing the users' rating data [1]. However, it still suffers from some drawbacks such as cold-start, sparsity and scalability problems.

Generally speaking, people tend to accept recommendations from familiar or trusted persons [2], and their opinion on a certain thing can greatly affect their friends' choice. With the development of online social networks, it becomes easier to collect and utilize the users' social relationship data. Some e-commerce companies examine how to leverage social relationships to improve the customers' purchase decision making so as to increase sales [3]. Researchers introduce a variety of social information to solve sparsity problems [4] and achieve better results. Modeling trust is mostly through the users' relationships. On the other hand, the users' social importance is another critical feature and can be used. User influence plays an important role in product marketing. He and Chu proposed a social recommendation system using user influence as a factor, and their experimental results on a dataset collected from www.yelp.com proved its superiority over collaborative filtering [5]. But such rec-

James J. (Jong Hyuk) Park et al. (eds.), *Future Information Technology*,
Lecture Notes in Electrical Engineering 309,
DOI: 10.1007/978-3-642-55038-6_13, © Springer-Verlag Berlin Heidelberg 2014

ommendation systems based on user influence are difficult to provide consistent and stable answers for active user's constantly changing query. Based on the above background, this work studies personalized recommendation on social media, such as www.yelp.com and www.dianping.com, which consist of the users' rating data and relationships. We propose a novel recommendation algorithm based on modeling the user's social trust and influence, making use of similarity, trust and user influence to predict the ratings. The experimental data set is crawled from www.dianping.com which allows people to follow others without the prior reciprocal agreement, so that the users' relationship forms a directed social network as Twitter.

2 Basic Model and Integrated Algorithm

In this section we provide the measurement model of similarity, trust and user influence, and outline the integrated recommendation algorithm.

The similarity of one's behaviors can be used to measure the similarity of one's interest. Existing metrics such as Pearson correlation coefficient and Cosin-based similarity are widely used to measure the similarities of users. Similarity based on Person Correlation coefficient is formulated as follows.

$$\mathrm{Sim}(m,n) = \frac{\sum_{i \in I_m \cap I_n}(R_{m,i}-\bar{R}_m)(R_{n,i}-\bar{R}_n)}{\sqrt{\sum_{i \in I_m \cap I_n}(R_{m,i}-\bar{R}_m)}^2 \sqrt{\sum_{i \in I_m \cap I_n}(R_{n,i}-\bar{R}_n)}^2} \tag{1}$$

where $\mathrm{I_m}$ $(\mathrm{I_n})$ is the set of items rated by user *m (n)*, $\mathrm{I_m} \cap \mathrm{I_n}$ are items rated by both *m* and *n*, $\overline{\mathrm{R}}_\mathrm{m}(\overline{\mathrm{R}}_\mathrm{n})$ indicates the average scores of *m* and *n* on all items of $\mathrm{I_m}(\mathrm{I_n})$, $\mathrm{R_{m,i}}(\mathrm{R_{n,i}})$ represents user *m*'s (*n*'s) rating on an item i.

A higher value for Sim(*m, n*) in Eq. (1) indicates a higher similarity in the preferences of two users. Using the preference similarity between user m and other users, we can predict the rating of user m on an item *i* as follows.

$$\mathrm{SimPre}(m,i) = \bar{R}_m + \frac{\sum_{u \in KNN_m} Sim(m,u)(R_{u,i}-\bar{R}_u)}{\sum_{u \in KNN_m}|Sim(m,u)|} \tag{2}$$

where $\mathrm{KNN_m}$ represents a user *m*'s top-k similar users who have rated item *i*, and $\mathrm{Sim(m,u)}$ represents the similarity between *m* and *u* respectively.

User influence is an important feature in direct social networks such as microblog network [6]. Number of followers is the most intuitive criteria because it has a positive correlation with user's ability to disseminate information. We propose an improved model to measure user influence according to the characteristics of the dataset. We model it with the nodes representing the users, and the directed edges representing the follow relation.

Figure 1 shows a graph that represents the user *U*'s relationship network. User *U*'s relationship network can be converted into a tree where *U*'s depth is 0. User *U*'s influence value is 3 when using the number of direct followers because *U* has three direct followers: *A*, *B* and *C*. However, indirect influence is also important, and the nodes whose depth is less than or equal to two are generally considered as the most valuable information [6].

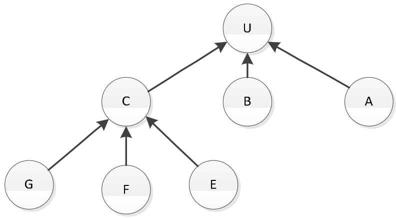

Fig. 1. A an example of user U's relationship network

If the common rated items of user U and his/her direct follower C occupy a relative high proportion of C's historical data, we consider U's influence on C is relatively high, and so is C's contribution to U's influence. The value of the contribution of each direct follower can be calculated as follows.

$$\mathrm{Contribution(u,c)} = F_c \frac{Common(u,c)}{RateNum(c)} \tag{3}$$

where c is u's direct follower and $\mathrm{Contribution(u,c)}$ represents c's contribution to u's influence. F_c represents the number of user c's direct followers, $\mathrm{Common(u,c)}$ stands for the number of common items rated by u and c, and $RateNum(c)$ indicates the number of items rated by c.

Combined with a number of direct followers and their contribution value, we can measure the user influence by Eq. (4).

$$W_i = F_i + \sum_{j=1}^{F_i} F_j \frac{Common(i,j)}{RateNum(j)} \tag{4}$$

where F_i represents the number of user i's direct followers. Then we predict the user m's rating on an item i based on user influence:

$$\mathrm{InfluPre}(m,i) = \bar{R}_m + \frac{\sum_{u \in KNN_i} W_u (R_{u,i} - \bar{R}_u)}{\sum_{u \in KNN_i} W_u} \tag{5}$$

where $\mathrm{KNN_i}$ represents the top-k influential users who have rated item i.

When buying products, people usually view others' evaluation, and consult their friends or professional users. Usually people tend to accept recommendations from people with close relationships. In this study, we try to build a trust measurement model based on the user's social network, and introduce a corresponding rating prediction algorithm.The structural dimension and the behavioral dimension are two ways to measure the social trust. The latter is used in this paper by following the so-called "six degrees of separation", a social theory that suggests everyone is considered to be six or fewer steps away.

The same as the models in social network analysis, we use the nodes to represent users and the directed edges to represent the follow relations respectively. Denote G (N, E) as the social networks where N represents the set of nodes, and E represents a collection of directed edges. If A follows B, then A is called follower, and B is to be followed. If two nodes follow each other, we consider them "friends".

Figure 2 shows an example of user *A*'s local social graph. We will process it in two steps. First, all users followed by *A*, direct and indirect, are rearranged to form a concentric circle. The number of the circle's hierarchy is less than six with the first layer nodes that represent the users followed directly by *A* and the second layer nodes that represent the users followed by the first layer' nodes and so on. Second, the following processing is done in order to save the shortest paths between *A* and other nodes only: (1) delete the directed edges between the nodes in the same layer; (2) remove the directed edges from the nodes in the outer layer to the nodes in the inner layer, but the links between the nodes who follow each other and are in the adjacent layers are excluded, such as the links between *I* and *J*; (3) the nodes' hierarchical number on the paths from *A* to a target node increases one by one.

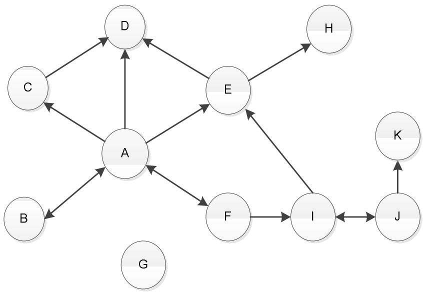

Fig. 2. An example of user *A*'s local social graph

The layer's maximum number is six. If any users are not in the processed concentric circle model of *A*'s social network, we set *A*'s trust on them is 0. For user *N* in the *A*'s concentric circle model, social trust value can be calculated as follows.

$$\mathrm{Trust(A,N)} = \frac{\mathbf{1}}{\boldsymbol{n+1}} \quad (0 < n \leq 6) \tag{6}$$

where *n* represents the layer number of *N*. There is one kind of situation that needs to be paid attention. If *A*'s friend is in the first layer, that is, they follow each other directly, we set the value of trust between them as 1.

The social trust measurement model conforms to the principle of sociology, and it is possible to select the number of layers according to the application scenario and computational complexity. As many trust-aware recommendation systems do, we use the following equation to predict the user m's rating on an item *i*.

$$\mathrm{TrustPre}(m,i) = (\theta\ \mathrm{SimPre\ (u,i)} + \varepsilon(\bar{R}_m + \frac{\sum_{u \in KNN_m} \mathrm{Trust}(m,u)(R_{u,i} - \bar{R}_u)}{\sum_{u \in KNN_m} |\mathrm{trust}(m,u)|})\ ) \tag{7}$$

where $\bar{R}_m$ and $\bar{R}_u$ denote the average rating of users m and u, $\mathrm{Trust}(m,u)$ represents the social trust value of m to u, and KNN_m represents the top-k trusted users who have rated item *i*.

After establishing the measurement models of social trust, user influence and introducing the rating prediction method, we describe the proposed algorithm in Table 1. The algorithm makes use of multiple information sources which provide more reference information, make the rating prediction independently and finally combine each result in a linear way. The algorithm degenerates to the user-based collaborative filtering if we only consider the user similarity.

Table 1. Social recommendation algorithm based on social trust and influence

Algorithm1: **Social recommendation algorithm**
Input: F (target user' social network), R (user-rating matrix), u (user), i (item), d (number of layers applied), α (social trust weight), β (user influence weight)
Output: $R(u, i)$ (user u's predicted rating on item i)
Step 1: set the value of d used to compute trust
Step 2: make the rating prediction based on social trust and user influence respectively
Step 3: calculate $R_{u,i}$ $=\alpha\,\mathrm{TrustPre}(u,i)\ +\ \beta\,\mathrm{InlfuPre}(u_m,i)$, and return $R_{u,i}$

3 Experiment Result and Analysis

In this section, we introduce an experiment based on real datasets to compare the performance of different algorithms. The experimental data set was collected from www.dianping.com, where the registered users can rate shops, restaurants, hotels or other services, and view others' comments, and the ratings range from 1 to 5. Users usually follow influential people or interested people, and their social relationships form a directed social network.

We crawled the rating data of hotels and the relationship data of users. The original dataset consists of 92,290 ratings from 45,109 users on 3,383 hotels. After pre-processing, those with numbers of total 5 or less rating are filtered out, the final dataset consists 36,106 ratings from 4,689 users on 2,608 hotels. We use the mean absolute error (MAE) as the accuracy metric to compare the performance of the proposed algorithm with collaborative filtering and other different algorithms. MAE is defined as follows.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |r_{u,i} - \hat{r}_{u,i}|}{n} \tag{8}$$

where n is the number of total items rated by the users, $r_{u,i}$ represents user u's predicted rating on an item I, while $\hat{r}_{u,i}$ denotes u's actual rating on i.

Exeriments were performed on different algorithms: the proposed algorithm, user-based collaborative filtering, trust-based and influence-based recommendation. The cross-validation methods were chosen and the mean value was taken as the final result. We conducted the experiments using the proposed social recommendation algorithm by adjusting parameters of social trust and user influence. Fig. 3 shows the result, where "9,1" means that the values of α, β are set to 0.9, 0.1 accordingly. Top-K is the top number of recommendations. The result shows that the prediction accuracy and stability of the algorithm is increased as the proportion of user influence grows in a certain range.

The optimum parameters of the proposed algorithm we got in this experiment is α=0.6, β=0.4. Fig.4 shows that the experimental results of different algorithms with

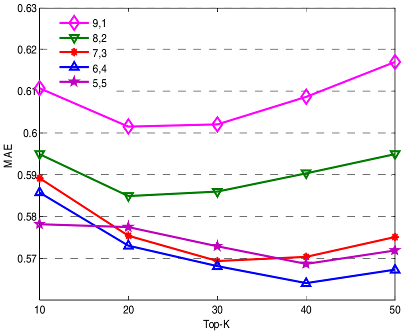

Fig. 3. Experimental results with adjusting parameters of trust and influence

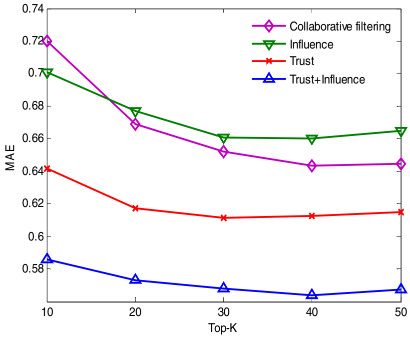

Fig. 4. Experimental results of different algorithms

the optimal parameters. If considering only a single factor, the trust-based algorithm can achieve better results when *K* is less than 20, and influence is not the better algorithm than the other two. However, our proposed algorithm combining both the social trust and user influence can get the best prediction accuracy.

4 Conclusion

In this paper, we proposed a new integrated recommendation algorithm combining both the trust and user influence. We modeled the user's social network using social network analysis method, and built a measurement model of trust and user influence. The experimental results showed that the proposed algorithm could achieve higher prediction accuracy, while reduce the impact of rating data sparsity, and consequently enhance recommendation accuracy and stability. Our study suggests that the proposed method has largely improved the quality of conventional collaborative filtering, and its quality could be further improved by means of designing more effective integrating schemes, which will be our future work.

Acknowledgment. This work is supported by the National Natural Science Foundation under Grant No. 61074315, and the Natural Science Foundation of Ningxia under Grant No. NZ12212.

References

1. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Communications of the ACM 35, 61–70 (1992)
2. Kim, H.W., Xu, Y., Gupta, S.: Which is more important in Internet shopping, perceived price or trust? Electronic Commerce Research and Applications 11(3), 241–252 (2012)
3. Kim, Y.A., Srivastava, J.: Impact of social influence in e-commerce decision making. In: Proc. 9th International Conference on Electronic Commerce, pp. 293–302 (2007)

4. Yang, B., Zhao, P., Ping, S., Huang, J.: Improving the Recommendation of Collaborative Filtering by Fusing Trust Network. In: Proc. 8th International Conference on Computational Intelligence and Security, pp. 195–199 (2012)
5. He, J., Chu, W.W.: A social network-based recommender system (SNRS). Springer US (2010)
6. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.: Measuring user influence in twitter: the million follower fallacy. In: Proc. International Conference on Weblogs and Social Media, pp. 10–17 (2010)