A Trust-based Friend Recommendation Strategy in Microblog

Shan Cui[1, a *], Ruifang Liu[1,b]

[1]Beijing University of Posts & Telecommunications, Beijing, China

[a]cuishan_bupt@163.com, [b]lrf@bupt.edu.cn

Keywords: Trust relationship; recommendation; personalization; online social networks

Abstract. With the increasing popularity of online social networks (OSNs), it is becoming a challenge to recommend proper friends for a target user. Recommendation system based on trust is one of the approaches having been developed in recent years. In this paper, we introduce a novel personalized recommendation strategy which is based on trust relationship. The experimental results show the strategy can more effectively use implicit and explicit connections between the OSNs users.

Introduction

With the rapid development of the Internet in recent years, online social networks (OSNs) provide platforms for hundreds of millions of people to share information and express opinions. As a result, numerous data emerge on blogs, microblogs, and other types of social networks. On the one hand, the explosion of information increases the difficulty to distinguish useful content and find user whom we are interested in. On the other hand, users would like to establish connections with more people who have something in common.

In order to solve the contradiction, recommendation systems, which can filter out the noisy effectively, are applied to recommend friends in OSNs. Most traditional recommendation systems are based on the similarity of content [1]. As a typical traditional recommendation system, collaborative filtering (CF) techniques recommend items bought or browsed by users who have similar tastes as you [2,3]. In this method, we manage to find patterns from users' behavior and preferences as users' feature. However, CF techniques ignore the fact that user values his friend's recommendation more while his friend and a stranger both have same tastes with him [4].

During the past years, recommendations based on relationships has got great development [5]. This type of recommendation tends to assume that user's friends have more influence on them than people without any links. In that case, social relationships in real life are simulated well [6]. However, some people add friends in real life to their followee list which means they don't follow friends for common interest. This kind of recommendation systems lead to meaningless users match by emphasizing relationship without similarity of content or interest.

In this paper, we proposed a model to evaluate whether a user will accept an item or not by utilizing both users' content preference and existing relationships. Assuming that relationships in OSNs composed a directed graph. In that graph, each node represents a user, and a directed edge from node A to node B represents A is a follower of B. Instead of finding the shortest path between two users, we utilize two layers of the graph to judge what kind of relationships exists between two specified users.

Related Work

Most recommendation system tends to recommend users who have similar tastes to a specify user. Similarity is defined by attribute such as age, sex, city or interests in most cases [7]. Although this kind of recommendation systems is easy to use, its application is limited. Sometimes, completed attribute of users is difficult to acquire. Besides, wrong personal data may lead to many improper recommendation.

Recommendation systems based on model used to train a valid model in machine learning ways and predict unknown result with this model. This method can utilize all kinds of features well and get better results in recommendation. Each user was assumed as an individual in this method. Relationships between users are ignored, nevertheless, they're more important than what we imagine.

Recommendation methods based on trust are more and more popular with the development of OSNs in recent years. Some existing works point out that trust relationships in OSNs can be more valid in recommendation systems. Ma et al. [8] propose an access recommendation algorithm based on trust relationship which constraint the objective function with trust relationship matrix. The purpose is achieved by characterizing feature vectors of target users close to their friends' in the process of learning. Jamali et al. [9] propose a method that integrate trust relationship with random walk model. Most recommendation algorithms assume that all trust relationships are homogeneous. Yeung et al. [10] deal with how trust relationships among users in Epinions effect final scores. They found that not every pair of users who established trust relationship grade a product close. It means that trust relationships are diversity. Each user owns a friend set which is used to maintain his trust relationship, in which all users have been trusted more or less by the owner. After every transaction, a trust value judged from the quality of the item or service should be rated towards another user, which will take effect in the next transaction, and work as an effective mechanism of trust feedback from user side. Finally, users could also check the reputation of other users through calculation of global trust, which gives a way to known expressions about someone from others [11].

Model Building

We regard friendship in OSNs as trust relationship. For example, if A follows B, in other words, A is a follower of B, then we assume that A trusts B, not vice versa. What's more, as is shown in Fig. 1, if A likes B and B likes C, then A trusts C to some extent. If A dislikes B and B likes C, then A distrust C. If A likes B and B dislikes C, then we assume A distrust C. This two-layer structure represents the spread of trust relationship. We take relationships in OSNs for directed edges in social graph. If A follows B, then the edge direct from A to B. A central concern of our research work is to judge whether an edge can generated or not. Any pair of nodes can be calculated a score that indicates the probability for new directed edge. We consider pairs of users whose scores beyond the threshold value that was defined before are probable new friends in the future.

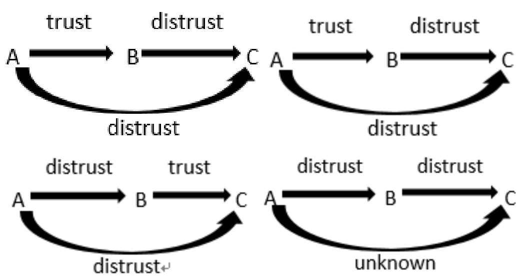

Figure 1. Propagation of trust relationship

In order to improve accuracy of recommendation results, we utilize explicit and implicit users' behavior and features. In microblog, users' preference expressed by not only behavior but also content of tweet and other text what are called item. Although recommendations are between users, attitude towards item plays an important role. The personalized recommendation strategy based on trust relationship in the paper can be divided into four parts.

The score of trust between two users is given by the following formula:

$$R_{ij} = w_{behavior}Behavior_{ij} + w_{trust}Interaction_{ij} + w_{links}Trust_{ij} + w_{similarity}Similarity_{ij} \quad (1)$$

where $w_{behavior}+w_{trust}+w_{links}+w_{similarity} = 1$.

As is shown in Eq. 1, the score of trust is a weighted combination of four factors. The first factor corresponds to the explicit behavior between users. The second factor corresponds to user's attitude towards items posted by target user. The third factor corresponds to relationships between two users. The fourth factor corresponds to the similarity between two users.

Explicit Behavior between Users

The system may recommend some similarity users calculated by specific algorithm to a specify user. As not every user is suited to the algorithm, user can choose accept or reject recommended users. There are plenty records of explicit behavior between users. If user A accepted user B, it won't recommend B to A anymore normally. In the other situation, that A would be unlikely to follow B if A has rejected B before. As a result, if B is recommended to A again, A's previous reaction has a lot of reference value to predict unknown result this time.

Positive recommendations can be used from the members in order to connect to new people (in social networking sites), subscribe to new blogs (in the blogosphere), share resources (in social bookmarking applications), etc. On the other hand, in the case of negative recommendations, the model in essence generates a list of untrustworthy users. This personalized blacklist can be exploited by the recommender system in order to alert users when content items are published from such untrustworthy users and discourage them from linking or browsing such content, or filter it out from their content feed [12]. Users' action of acceptance or rejection express their trust preference which performs an important function in recommendation system. We add users accepted by A to A's likes list which contains users trustworthy, and add users rejected by A to A's dislikes list which contains users untrusted. The factor can be modeled as a binary decision variable taking values 1 or -1.

Interaction between Users

People tend to communicate with trustworthy friends, while running a mile from untrusted users. On this occasion, the more users interact, the more they trust. Not too many interaction between users in microblog service, since it's a platform to share information in forms of texts, pictures and videos. In that case, user's action towards item represents his attitude of trust. When repost or comment one's tweet, level of trust increased a bit. Other types of action, such as at and send a private message, also means trust relationship is becoming more solid. Due to different activeness, actions should not be compared by count. It is common sense that one interaction from user who interaction twice is more significant than ten interactions from user who interaction a hundred times. Percentage is much more proper than number of times. The factor has been assumed to lie within the [0, 1] range and is defined as follows:

$$Interaction_{ij} = \frac{number\ of\ actions(i \rightarrow j)}{\sum_{k\ in\ i's\ action\ list} number of\ \ actions(i \rightarrow k)} \tag{2}$$

where a value close to 1 indicates that the target user is trustworthy to the evaluator user.

Level of Trust

In the paper, trust propagation calculates the reliability of a trusted path. There are three cases that should be taken into consideration. Firstly, if user A follows user B and user B follows user C, then we consider user A trust user C to some extent. Secondly, if user A follows user B and user C is in user B's blacklist, then user A doesn't trust user C either. Thirdly, if user B follows user C and user B is in user A's blacklist, then user A doesn't trust user C. It is shown that trust (distrust) relationship can be transferred. The situation that user B is in user A's blacklist and user C is in user B's blacklist, left untreated, since too many other factors to be considered.

Basically, the more common friends two users have, the more trust between them. Followers are not considered, since they are passive results for the specific user. Common friends are defined as the intersection of users' followee list, as in the following:

$$\text{Common friends(i,j)} = (\text{i}'\text{s followee list}) \cap (\text{j}'\text{s followee list}) \tag{3}$$

We define common attraction as an essential measure of trust. Common attraction is expressed as the intersection of users' followee list divide the union of users' followee list instead of the number of common friends. The quotation is defined as follows:

$$\text{Common attraction(i, j)} = \frac{(\text{i}'\text{s followee list}) \cap (\text{j}'\text{s followee list})}{(\text{i}'\text{s followee list}) \cup (\text{j}'\text{s followee list})} \tag{4}$$

All users in followee list and users accepted by the target user are classified as friends which represents trust relationship. Users rejected by the target user are classified into blacklist which means no trust relationship exists. For a specify user, we first acquire his friend list which known as the first layer of friends. In the meantime, blacklist can be acquired as the first layer of dislikes. Each user in the first layer of friends has their own friend list. All of those friend list compose the second layer of friends. Likewise, each user in the first layer of friends has their blacklist, and each user in the first layer of dislikes has their friend list. These two parts compose the second layer of dislikes. As is shown in the Fig. 2, the black dot represents the target user while triangle means that the user is in friend list and square means the user is in blacklist. All triangle elements above the diagonal line in the second layer compose friend list, and all squares below the diagonal line means unknown. The others in the second layer compose the blacklist. The next thing to do is to calculate how many times that the evaluator user appear in the second layer of friends what we call indirect friend list and the second layer of dislikes what we call indirect blacklist.

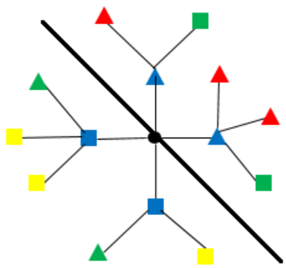

Figure 2. Trust relationship in two layers

According to links between users, relationship score is given by

$$\text{trust}(i,j) = \frac{\text{count}(j)\ \text{in indirect friend list}}{\text{total indirect friends}} - \frac{\text{count}(j)\ \text{in indirect blacklist}}{\text{total members indirect blacklist}}. \tag{5}$$

Personalized Similarity

Trust relationships are diversity, so personalized factor is needed. Text similarity is taken into consideration. People tend to trust users who have similar opinions or interests with them. In microblog, most of the time, users express their thought via text due to its originality and intuition. Text similarity can be obtained from tweet, tags or individual signatures. For a specify dataset, keywords are needed since comparison between two pieces of text is poor operation. Methods of acquiring keywords won't go into details in this paper since the dataset contains keywords already. We use cosine similarity which is widely used in collaborative filtering to measure the similarity between users. Text similarity is given by

$$\text{Similarity}(i,j) = \frac{\vec{i}\cdot\vec{j}}{\|\vec{i}\|\|\vec{j}\|}. \tag{6}$$

The similarity takes any value in the [0, 1] range.

Experiments

In this section, we evaluate the performance of our algorithm with experiments in KDDCup2012's data set. For experimentation, we used 2972 records consist 100 users. All data needed are listed in Table 1. Since keywords and tags in this data set are encrypted as numbers, we only need to judge whether same numbers are exist in two keywords (tags) list. Similarity is calculated by cosine similarity with keywords' weight.

TABLE I. DATA COLLECTED OF USERS

Part	Data Collected
Explicit behavior	Recommendation History
Trust towards item	Number of actions
Links	Followee list, Friend list, Blacklist
Personalized similarity	Keywords, Tags

Table 2 shows the result of our model in predicting a friends list of a target user. Since history of recommendation is included, action of reject may lead to negative score. The precision and recall varies according to the value of the threshold. Lower threshold means positive trust scores play a larger role, while higher threshold means negative trust scores play a larger role.

TABLE II. EXPERIEMNT RESULT

Threshold	Right Precision	Right Recall
-1	0.295	0.839
-0.5	0.296	0.802
0	0.317	0.684
0.1	0.398	0.494

Conclusion

In this paper, we propose a personalized trust-based recommendation method, where we calculate scores which combine trust relationship and text similarity instead of extracting many features. This method can help people make proper decisions and easy to use with less training time. Through analysis of experiments in a real OSNs data set, we find that recommendation based on trust relationship in virtual community can be more reasonable. In the future work, we will consider timestamp information and add more negative connections to make more accurate decisions.

References

[1] Michael J. Pazzani, Daniel Billsus. Content-based recommendation systems. Lecture Notes in Computer Science, Volume 4321, pp 325-341, 2007.

[2] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. Item-based collaborative filtering recommendation algorithms. Proceedings of the 10th international conference on World Wide Web, pages 285-295, 2001.

[3] Xiongcai Cai, Michael Bain, et al. Collaborative filtering for people to people recommendation in social networks. Lecture Notes in Computer Science, vol.6464, pp 476-485, 2010.

[4] Jing Wang, Jian Yin. Combining user-based and item-based collaborative filtering techniques to improve recommendation diversity. 6th International Conference on Biomedical Engineering and Informatics, 2013.

[5] Yuxiao Dong, Jie Tang, Sen Wu, Jilei Tian, et al. Link prediction and recommendation across heterogeneous scoial networks. Data Mining (ICDM), IEEE 12th International Conference, pp 181-190, 2012.

[6] Xiwang Yang, Harald Steck, Yong Liu. Circle-based recommendation in online socila networks. Proceeding of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 1267-1275, 2012.

[7] Fan Tang, Bofeng Zhang, Jianxing Zheng, Yajun Gu. Friend recommendation based on the similarity of micro-blog user model. 2013.

[8] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, Irwin King. Recommender systems with social regularization. Proceeding of the fourth ACM international conference on Web search and data mining, pp 287-296, 2011.

[9] Mohsen Jamali, Martin Ester. Trustwalker:A random walk model for combining trust-based and item-based recommendation. Proceeding of the 15th ACM SIGKDD international conference on Knwoledge discovery and data mining, pp 397-406, 2009.

[10] Ching-man Au Yeung, Tomoharu Iwata. Strength of social influence in trust networks in product review sites. Proceedings of the fourth ACM international conference on Web search and data mining, pp495-504, 2011.

[11] Wu Wenjuan, Lu Zhubing. A personalized recommendation strategy based on trusted social community. 10th International Conference on Computer Science & Education, pp 496-499, 2015.

[12] Eirinaki, M, Louta, M.D., Varlamis, I. A trust-aware system for personalized user recommendations in social networks. IEEE Transactions on Systems, volume 44, pp 409-421, 2014.