International Journal of Control and Automation
Vol.8, No.11 (2015), pp.21-32
http://dx.doi.org/10.14257/ijca.2015.8.11.03

Social Friend Interest Similarity in Microblog and its Implication

Yan Yu[1], Lingfei Mo[2] and Jiasheng Zhou[1]

[1]*Computer Science Department, Southeast University Chengxian College, Nanjing 210088, China*
[2]*School of Instrument Science and Engineering, Southeast University, Nanjing 210009, China*
yuyanyuyan2004@126.com

Abstract

Recently microblog has become a popular Internet service. Yet, the abundance of micrblog floods users with huge volumes of information and poses a great challenge in terms of information overload. Recommender systems aim to alleviate information overload over microbog users by presenting the most attractive and relevant content. Many recommendation approaches have been proposed by leveraging the social relationships in microblog. To prove the feasibility of users' social relationships as the bases of recommendation in microblog, we examine the correlations of social relationships and user interest similarities in microblog. Using real-world data set, we find that social relationship indicates positive connection with user interest similarity in microblog, but the positive connection is not strong. We also observe strong positive correlation between reciprocal social relationship and user interest similarity in microblog. As to users' interaction, we find that the number of interaction between two users in social relation is a strong signal that controls users' interest similarity. We then apply these findings to recommendation application in microblog to improve the accuracy of recommendation. The experiment on real-world data set shows that our findings are useful for recommendation.

Keywords: *Interest similarity, social relationship, microblog, recommender systems*

1. Introduction

Recently, microblog has become a popular Internet service. Millions of users are active daily in microblog, creating rich information online that has not been available before. Yet, the abundance and popularity of microblog floods users with huge volumes of information and hence poses a great challenge in terms of information overload. Recommender systems aim to alleviate information overload over microblog users by presenting the most attractive and relevant content. Recommendations of content, tags, people and communities often use the personalization techniques adapted to the needs and interests of the individual user.

Users form an explicit social network by following other user in microblog. A user as a follower automatically receives the messages posted by the users he/she follows, known as followee or social friend. Many recommendation approaches in microblog have been proposed by leveraging the topological structure of formed social networks in microblog [1-5]. The basic assumption behind these approaches is based on the principle of homophily. Homophily shows the tendency of individuals to associate and bond with similar others. Individuals in homophilic relationships share common characteristics that make communication and relationship formation easier.

Some researchers have investigated the correlations of social relationships and users interest similarities in online social networks to provide fundamental support

ISSN: 2005-4297 IJCA

to the research of friend-based social recommendation problem in online social networks [6-8]. Zielger in [6] conducted analysis on users observed in two online social networks which allow their users to express their personal preference, such as the members they trust and the products they appreciate. They found strong indication towards positive correlation between trust relationship and user interest similarity. Lee and Brusilovsky in [7] found that users connected by a self-defined relationship of trust have more common information items and meta-data than users with no connection. Ma in [8] conducted an analysis on friend-based social network. They found that social relations generally cannot represent user interest similarities in friend-based social network.

Although researchers have done some work previously, many research questions are still left open and need to be further explored. First of all, social network formed by users in microblog differs substantially from other online social networks, such as ones in Facebook or LinkedIn, where social relationships can only be established with the consent of both to-be connected users. In contrast, a social relationship in microblog is asymmetric. In other words, a user can follow a followee without the followee’s consent. The asymmetry of social ties in microblog has made microblog social networks called hybrid networks [9]. They are hybrid because users create social relationships not only for communicating with friends or acquaintances but also for getting information on particular subjects [10-11]. Kwak [12] revealed that 77.9% of users’ social relationships are not reciprocal in Twitter. Furthermore, 67.6% of users don’t contain any reciprocal social relationship, which means that the majority of users in microblog simply look for interesting information rather than keeping in touch with their friends [12]. It is hence important to study the correlations between social relationships and user interest similarities in the context of microblog to provide fundamental support to the research of recommendation in microblog. The research questions we explore in this paper are: Does social relationship indicate positive connection with user interest similarity in microblog? Does reciprocal social relationship indicate positive connection with user interest similarity in microblog?

Secondly, besides following other users, a user can also interact with other users in microblog. If a user wants to notify another user about his/her message, he/she would use an '@' to notify the other user. A user can repost a message and append some comments to share it with his/her followers. A user can add some comments to a message. The number of interaction between two users is often considered the strength of their relationship. Researcher has found that pairs of individuals with strong tie exhibit greater similarity than those with weak ties in online social networks [13]. In microblog, fantastic football fans may repost football stars many more times than other persons they follow. Hence, a natural research question we can explore is: Is the number of interaction between two users in social relationship a strong signal that controls user interest similarity in microblog?

Our main contributions in this paper are two fold. First, we conduct several experiments on a real-world data set to address the aforementioned research questions. The major findings from the first part are summarized as follow:

1. We find that social relationship indicates positive connection with user interest similarity in microblog.
2. We also observe strong positive correlation between reciprocal social relationship and user interest similarity in microblog, but the positive connection is not strong.
3. As to users' interaction, we find that the number of interaction between two users in social relationship is a strong signal that controls user interest similarity.

Secondly, we apply our findings to recommendation application in microblog to improve the accuracy of recommendation. The experiment on real-world data set shows that our findings are useful for recommendation which relies on user interest modeling in microblog.

The remainder of this paper is organized as follows. Section 2 introduces several related work in the literature. Section 3 conducts experiments on a real-world data set to investigate the correlations between social relations and user interest similarities in microblog. The implication of our findings for recommender systems in microblog are summarized in Section 4, followed by the conclusion and future work in Section 5.

2. Related Work

In this section, we review several related work, which mainly focused on user interest similarity analysis in online social networks.

Ziegler in [6] conducted analysis on users observed in two online social networks. These two online social networks allow their users to express their personal preferences, such as the members they trust and the products they appreciate. They argued that in order to provide meaningful results, trust must reflect user interest similarity to some extent because recommendations only make sense when obtained from like-minded people exhibiting similar taste. Their analysis concluded that they found strong indication towards positive correlation between trust and user interest similarity.

Lee and Brusilovsky in [7] argued that users’ self-defined networks of trust could be valuable to increase the quality of recommender systems. To prove the feasibility of users' self-defined relationships of trust as the bases of recommendation, they examined how similar interest of users connected by a self-defined relationship of trust are by using real life data collected from a social web system. They found that users connected by a self-defined relationship of trust have more common information items and meta-data than users with no connection. They also noticed similarity is largest for direct connections and decreases with the increase distance between users in the network of trust. Users involved in a reciprocal relationship exhibit significantly larger similarity than users in a unidirectional relationship.

Ma in [8] argued that trust is only one of many types of social relations in online social networks. Social friendships are quite different from trust relationships. The hypothesis in trust relationships may not be held in friend-based recommender systems. In order to provide fundamental support to the research of friend-based social recommendation problem, they conducted an analysis on the correlations between social friend relations and user interest similarities on two large friend communities extracted from real world recommender systems. They found that social friend relations generally cannot represent user interest similarities. They also noticed that in the social friend communities, a user’s similarities with his/her friends are very diverse. The similarities between a user and his/her friends are actually controlled by the network structure in the friend network.

However, as mentioned in Section 1, there are still many research problems need to be further investigated. In the next section, we provide an investigation on the study of correlations between social relations and user interest similarities in microblog.

3. Experiment Analysis on Social Friend Interest Similarity in Microblog

In this section, we conduct experiments to investigate the interest similarities of users connected by social relationships in microblog. We first describe the data set used in this paper. We then define interest similarity metric for evaluation. Last, we give detailed experimental analysis.

3.1. Data Set Description

The data set we use in this paper is from Tencent Weibo, which is a Chinese microblog website launched by Tencent in April, 2010. Tencent Weibo has become one of leading microblog platforms in China. Similar to Twitter, a user in Tencent Weibo can broadcast a short message up to 140 characters and follows other users on the website. Besides following other users, a user can also interact with other users in Tencent Weibo. A user can @, repost and comment on other users. The data set we use in this paper is from KDD Cup 2012 Track 1, which is a prediction task that involves predicting whether or not a user will follow a recommended user. The data set is a sampled snapshot of Tencent Weibo, including user profiles, social graph, interaction, and so on. In this paper, we only use social graph and interaction for our analysis. Table 1 shows the descriptive statistics.

Table 1. Data Summary of Tencent Weibo

Total number of users	2,320,895
Total number of social relationships	50,655,143
Total number of @	899,899
Total number of repost	8,790,544
Total number of comment	2,179,510

Without loss of generality while keeping our test meaningful and manageable, we randomly sample 81,319 users from the KDD Cup 2012 Track 1 data set as our target users to investigate their interest similarities with their social friends. In order to reduce noises, we require that each target user needs to have at least five claimed social relations.

3.2. Definition of Interest Similarity

Interest similarity we explore in this paper is based on the homophily principle of shared interest [9]. In microblog, shared interests can be represented by u→k←v, where user u and user v both follow user k. u and v sharing interests is surely one kind of similarity in microblog. In this paper, we use Jaccard Similarity Coefficient to define the similarity between two users u and v based on the followees they follow in common [5].

Mathematically, we can construct a directed graph G (V, E), where V represents a set of users in microblog and E represents a set of social relationships among these users. A directed edge <u, v> ∈E exists between user u and v if u follows v. The set of out-neighbors of user u is $\Gamma_+(u) = \{v \in V \mid \langle u,v\rangle \in E\}$, and the out-degree of u is $|\Gamma_+(u)|$, where $|\cdot|$ denotes the size of the set. Thus, the interest similarity between user u and user v is defined as:

$$sim(u,v) = \frac{|\Gamma_+(u) \cap \Gamma_+(v)|}{|\Gamma_+(u) \cup \Gamma_+(v)|} \tag{1}$$

3.3. Social vs. Non-Social Friends

The first analysis we perform is to understand the research question: Does following relationship indicate positive connection with user interest similarity? More specifically, we conduct the experiment as follows:

1. For each target user u, we calculate the average interest similarity between u and his/her social friends:

$$sim(u) = \frac{\sum_{v \in \Gamma_+(u)} sim(u,v)}{|\Gamma_+(u)|} \tag{2}$$

2. We also calculate the average interest similarity of target user u with non-social friends:

$$n\_sim(u) = \frac{\sum_{v \in R(u)} sim(u,v)}{|R(u)|} \quad (3)$$

where R(u) represents the list of randomly selected users whom user u does not follow. R(u) has the same size with $\Gamma_{+}(u)$, and $R(u) \cap \Gamma_{+}(u) = \emptyset$.

3. We then compare these two average interest similarities for each target user in great detail.

Figure 1 plots the correlations between average social friends' interest similarities and non-social friends' interest similarities of target users on date set. Every data point in Figure 1 represents a target user with the x-axis specifying average social friends' interest similarities and the y-axis indicating the related average non-social friends' interest similarities.

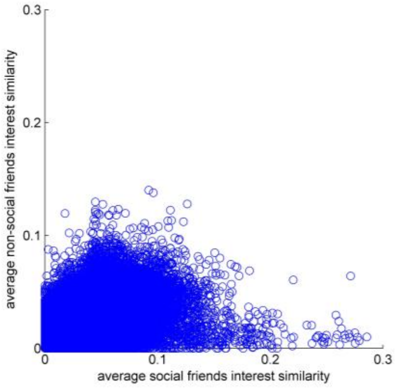


Figure 1. The Correlations between Social and Non-Social Friends Interest Similarity

We observe that the plot in Figure 1 exhibits biases towards lower-right region, which indicates that social relations has correlation with user interest similarity in microblog However, we also notice that the biases towards the lower-right region are not strong, which shows indication that the correlation between social relation and user interest similarity is not strong in microblog. In order to quantify the correlations between average social friends' interest similarities and average non-social friends' interest similarities, we measure the proportion of users whose average social friends' interest similarities are greater than their average non-social friends' interest similarities (i.e., sim(u)>n_sim(u)), which we call similarity ratio. Totally, similarity ratio is 61.8% of users whose average social friends' interest similarities are greater than their non_social friends' interest similarities in microblog. Table 2 provides an overview of similarity ratio compared to the ones from prior research in online social network. Table 2 shows the similarity ratios are 45.1% and 52.8% in Douban and Foursquare respectively, which indicate social relation is not correlated with user interest similarity in Douban and Foursquare. However, there are 82.9% of users whose social friends' interest similarity are greater than their non_social friends' interest similarity in Epinions, which indicates social relation has high correlation with user interest similarity in Epinions.

Table 2. Comparison of Online Social Networks' Similarity Ratio

Online Social Networks	Number of Users	Similarity Ratio

Douban [8]	129,490	45.1%
Foursquare [8]	16,748	52.8%
Epinions [8]	51,670	82.9%
Tencent Weibo	81,319	61.8%

Users tend to look for interesting information by following other uses in microblog [7]. So, social relation has correlations with user interest similarity. However, the correlation is not strong. We can understand this from three aspects. First of all, users trend to follow celebrities in microblog because these celebrities are well known to people and they are more likely to be reliable and stable information sources [14]. Consequently, even if two users don't have social relationship, they may both follow the same celebrity. Secondly, users play a dual role in microblog as they are both information sources and seekers [8]. What a user follows may not wholly similar with what he/she is followed. Lastly, there exist some spammers in microblog. One of the most common ways for spammers to gain popularity and get more spam targets is to follow a huge number of users and wait for them to follow back [15]. Hence, we cannot observe strong correlations between social following relationships and user interest similarity.

3.4. Unidirectional vs. Reciprocal Social Relations

Unlike some online social networks, a followed user has the option but not the requirement to similarly follow back. Thus, social relationships in microblog may be asymmetric. We distinguish two kinds of following relationships between two users in microblog: unidirectional and reciprocal. If user u follows user v, but v does not follow u, we call the social relationship between u and v as unidirectional and we call v is u's unidirectional social friends. If user u and user v both follow each other, we consider them reciprocal friends and we call v is u's reciprocal friend. Reciprocity is a source of social cohesion. When two individuals attend to one another, the bond is reinforced in each direction and both people will find the tie rewardings [16]. Hence, the second experiment we conduct is to compare the differences of interest similarities between unidirectional and reciprocal social relations. More specifically, for any target user who has reciprocal social relations, we conduct the experiments as follows:

1. For each target user u who has reciprocal social relations, we calculate the average interest similarity between u and his/her unidirectional social friends:

$$uni\_sim(u) = \frac{\sum_{v \in Uni(u)} sim(u,v)}{|Uni(u)|} \tag{4}$$

where Uni(u) represents the list of unidirectional social friends of user u follows.

2. For each target user u who has reciprocal following relations, we also calculate the average interest similarity between u and his/her reciprocal social friends:

$$reci\_sim(u) = \frac{\sum_{v \in \mathrm{Re}ci(u)} sim(u,v)}{|\mathrm{Re}ci(u)|} \tag{5}$$

where Reci(u) represents the list of reciprocal social friends of user u .

3. We then compare the values between uni_sim(u) and reci_sim(u) for each user u.

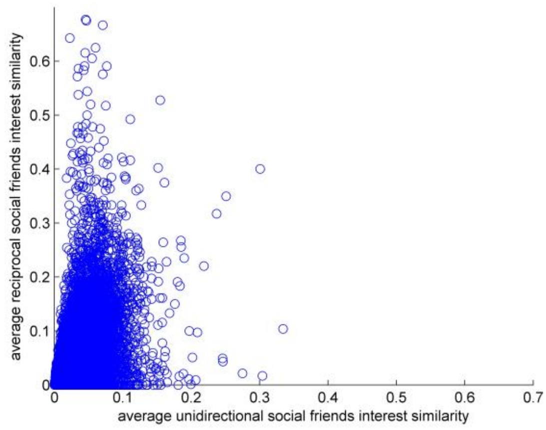


Figure 2. The Correlations between Unidirectional and Reciprocal Social Friends Interest Similarity

Figure 2 plot the correlations between average unidirectional social friends' interest similarity and reciprocal social friends' interest similarity on data set. Every data point in Figure 2 represents a target user with the x-axis specifying average unidirectional social friends' interest similarities and the y-axis indicating the related average reciprocal social friends' interest similarities. We notice that the plot in Figure 2 exhibits strong biases towards the upper-left region, which shows a strong indication that reciprocal social relationship has high correlation with user interest similarity. Users involved in a reciprocal social relationship exhibit significantly larger similarity than users in a unidirectional social relationship.

3.5. Number of Interaction

The third experiment we conduct is to evaluate how the number of interaction between a pair of users in social relation can affect the interest similarity between these two users in microblog. More specifically, for any social relationship <u,v> between user u and user v in the data set, we calculate the peer similarity sim(u,v) mentioned in Equation 1. Then we count the number of @, repost and comment from u to v respectively. The aggregated results in Tencent Weibo are shown in Figure 3, 4 and 5 respectively.

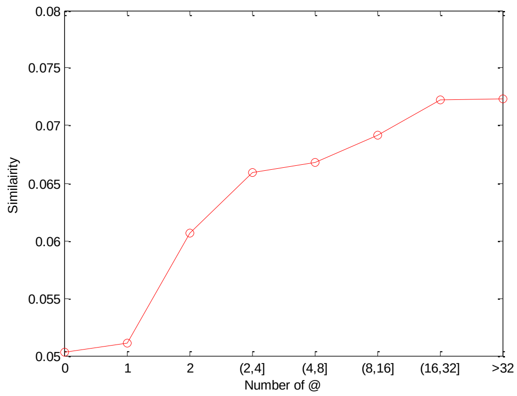


Figure 3. Similarities Conditioned on the Number of Interaction @

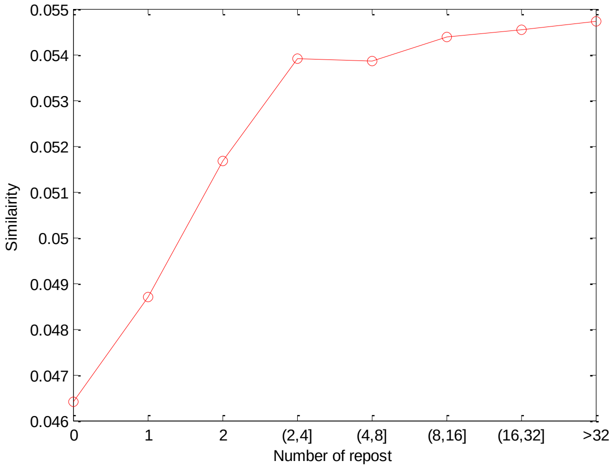


Figure 4. Similarities Conditioned on the Number of Interaction Repost

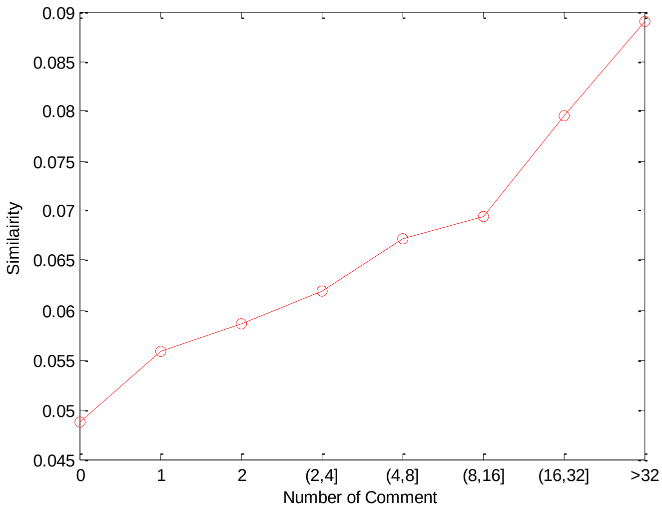


Figure 5. Similarities Conditioned on the Number Of Interaction Comment

In the x-axis of Figure 3, 4 and 5, we group the number of interaction into 8 categories, where "(4, 8]" indicates the number of interaction is greater than 4 but less or equal to 8. From Figure 3, 4 and 5, we can see that two users involved in a social relation are more similar if there are more interactions between them. One interpretation for this observation is that many interactions from a user may indicate the strong interest in the message posted by his/her social friend.

4. Implication for Item Recommendation in Microblog

From Section 3, we get some interesting observations. We believe these observations can be exploited by recommender systems and other applications, which rely on user interest modelings. In this section, we apply these observations to the item recommendation in microblog to improve the accuracy of recommendation. In this section, item recommendation in microblog is to recommend most attractive and relevant celebrities, organizations, or groups, for users to follow.

Recently, researchers found that by incorporating social relation information into the asymmetric factor models (AFM), the recommendation accuracy of tradition matrix factorization techniques can be significantly improved [17]. In matrix factorization, each user u is associated with a feature vector p_u of dimensionality d and each item i is associated with a feature vector q_i of dimensionality d. The result dot prodcut $p_u^T q_i$ represents the user u's overall interest in the item i. Hence, the estimated match value of user u and item i is . Matrix factorization models seek to approximate p_u and q_i by minimizing the

sum-of-squared-errors between r_{ui} and . In AFM, a user's feature p_u is represented by the features of his /her followees. Hence, in AFM, the estimated matching value of user u and item i is:

$$\hat{r}_{ui} = (\frac{1}{\sqrt{|\Gamma_+(u)|}} \sum_{j \in \Gamma_+(u)} q_j)^T q_i \tag{6}$$

In this section, we incorporate our findings from Section 3 into the AFM to improve the recommendation accuracy. In Section 3, we find that reciprocal social relationship has high correlation with user interest similarity and two users in a social relation are more similar if more interaction between them in microblog. Hence, we represent a user's feature p_u by the features of his/her social friends, reciprocal social friends and his/her social friends whom he/she interacts with. Thus, the estimated matching value of user u and item i is:

$$\begin{aligned} \hat{r}_{ui} = (&\frac{1}{\sqrt{|\Gamma_+(u)|}} \sum_{j \in \Gamma_+(u)} q_j \\ &+ \frac{1}{\sqrt{|\mathrm{Re}\,ci(u)|}} \sum_{j \in \mathrm{Re}\,ci(u)} q_j \\ &+ \frac{1}{\sqrt{|N(u)|}} \sum_{j \in A(u)} N(u,j) q_j)^T q_i \end{aligned} \tag{4}$$

where N(u) is the set of items user u interacts with. N(u,j) is the number of interactions user u has over his/her social friend j.

In this section, we conduct experiment to evaluate our method by utilizing the real-world data set from Tencent Weibo mentioned above. We make the snapshot of users' following information on October 11, 2011 and interactions between these users as a training set. We then use records of following history from 10/11/2011 to 30/11/2011 as the test set. We use precision to evaluate the accuracy of recommendation algorithms in this paper. Precision measures the average percentage of the overlap between a given recommendation list and the list of followees that are actually followed. Precision can be evaluated at different points in a ranked list of recommended users.

The result of comparison is shown in Figure 5. From the result in Figure 5, we observe that our method outperforms AFM method. Our method incorporates reciprocal social relations and interaction information to represent user's feature accurately.

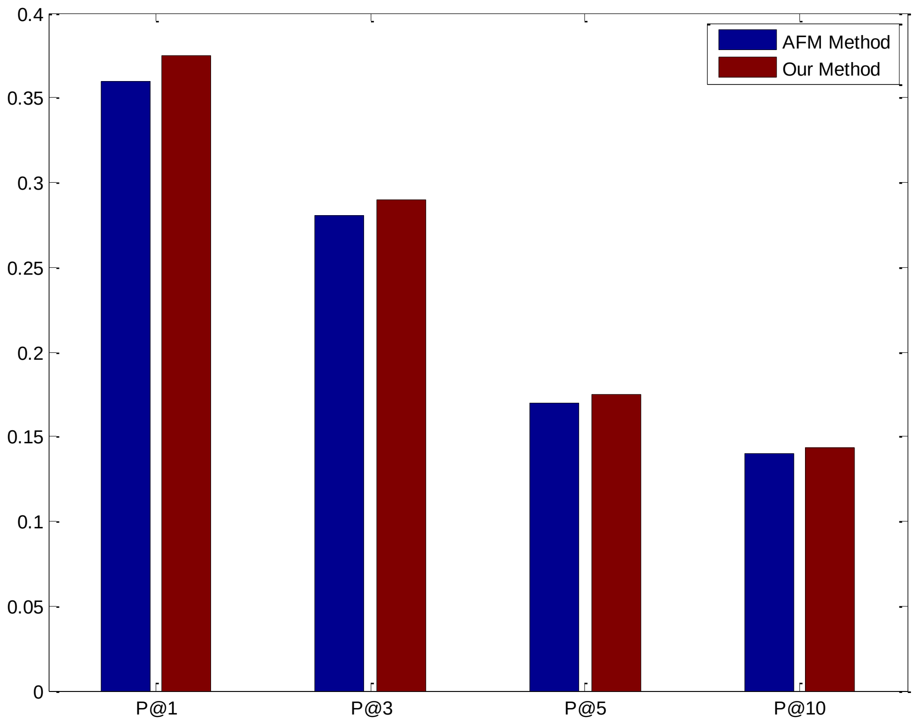


Figure 5. Accuracy Comparison of AMF Method and our Method

5. Conclusion and Future Work

To prove the feasibility of users' social relations as the bases of recommendation in microblog, we examine the correlations of social relations and user interest similarities in microblog. Using Tencent Weibo data set, we find that social relation indicates positive connection with user interest similarity in microblog, but the positive connection is not strong. We also observe strong positive correlation between reciprocal social relationship and user interest similarity in microblog. As to users' interaction, we find that the number of interaction between two users in social relation is a strong signal that controls user interest similarity. We then apply these findings to recommendation application in microblog to improve the accuracy of recommendation. The experiment on Tencent Weibo data set shows that our findings are useful for recommendation.

In our future studies we plan to include more information from microblog to estimate the user interest similarity and further improve the accuracy of item recommendation in microblog.

Acknowledgements

This research was supported by the Research on University's Natural Science Projects of Jiangsu Province (no. 14KJB520004),
the National Key Technology R&D Program (2013BAJ05B02) and Jiangsu's Qing Lan Project (11038).

References

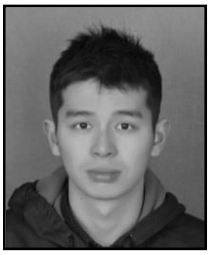
[1] J. Lin, K. Sugiyama, M-Y Kan, T-S Chua and G. J. F. Jones, "Addressing cold-start in app recommendation: latent user models constructed from Twitter followers", Proceedings of the 36th

International ACM SIGIR Conference on Research and Development in Information Retrieval, **(2013)** July 28- August 1; Dublin, Ireland.
[2] I-H Ting, P. S. Chang and S-L Wang" ,Understanding microblog users for social recommendation based on social networks analysis", Journal of Universal Computer Science, vol. 4, no. 18, **(2012)**.
[3] Y. Yamaguchi, T. Amagasa, H. Kitagawa and E. Kiciman, Editor, "Recommending fresh URLs using Twitter list", Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, **(2013)** July 8-11; Massachusetts, USA.
[4] Y. Yu and X. Wang, "Link prediction in directed network and its application in microblog", Mathematical Problems in Engineering, vol. 1, no. 1, **(2014)**.
[5] Y. Yu and R. G. Qiu, "Followee recommendation in microblog using matrix factorization model with structural regularization", The Scientific World Journal, vol. 1, no. 1, **(2014)**.
[6] C-N Ziegler and J. Golbeck, "Investigating interactions of trust and interest similarity", Decision Support System, vol. 2, no. 43, **(2007)**.
[7] D. Lee D and P. Brusilovsky and D. Jannach, Editor, "Does trust influence information similarity?", Proceedings of the ACM RecSys'09 Workshop on Recommender System & the Social Web, **(2009)** October 25-28; New York, USA.
[8] H. Ma and S. Geva, Editor, "On measuring social friend interest similarities in recommender system", Proceedings of the 37th International ACM SIGIR Conference on Research and Development in information Retrieval, **(2014)** July 6-11; Gold Coast, Australia.
[9] D. Yin, L. Hong, B. D Davison and B. Berendt, "Editor. Structural link analysis and prediction in microblogs", Proceedings of the 20th ACM International Conference on Information and Knowledge Management, **(2011)** October 24-28; Glasgow, United Kingdom.
[10] A. Java, X. Song, T. Finin, B. Tseng. H. Zhang, Editor. Why we twitter: understanding microblogging usage and communites, Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, **(2007)** August 12- 15; San Jose, USA.
[11] B. Krishnamurthy, P. Gill, M. Arlitt. C. Faloutsos, Editor. A few chirps about twitter. Proceedings of the First Workshop on Online Social Networks, **(2008)** August 17-22; Seattle, USA.
[12] H. Kwak, C. Lee, H. Park, S. Moon. M. Rappa, Editor. What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, **(2010)** April 26-30; Raleigh, USA.
[13] W. Christo, B. Bryce, S. Alessandra, K. P. Puttaswamy, B. Y Zhao. W. S. Preikschat, Editor. User interactions in social networks and their implications. Proceedings of the 4th ACM European Conference on Computer Systems, **(2009)** April 1-3; Nuremberg, Germany.
[14] M. Wilson and C. Nicholas. I. Soboroff, Editor. Topological analysis of an online social network for older adults. Proceeding of the 2008 ACM Workshop on Search in Social Media, **(2008)** October 26-30; Napa Valley, USA.
[15] Y. Zhou, K. Chen, L. Song, X. Yang, J. He. B. Werner, Editor. Feature analysis of spammers in social networks with active honeypots: a case study of Chinese microblogging networks. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, **(2012)** August 26-29; Istanbul, Turkey.
[16] S. A. Golder and S. Yardi. B. Werner, Editor. Structural predictors of tie formation in twitter transitivity and mutuality. 2010 IEEE Second International Conference on Social Computing, **(2010)** August 20-22; Minneapolis, USA.
[17] T. Ma, Y. Yang, L. Wang, B. Yuan. Recommending people to follow using asymmetric factor models with social graphs. Soft Computing in Industrial Applications Advances in Intelligent Systems and Computing, vol. 1, no. 223, **(2014)**.

Authors

Yan Yu, Female. Doctor. Computer Science Department, Southeast University Chengxian College, Nanjing 210088, China. Research Area: Recommender Systems.

Lingfei Mo, Male. Doctor. School of nstrument Science and Engineering, Southeast University, Nanjing 210009, China. Research Area: Artificial Intelligence.



Jiasheng Zhou, Male. Computer Science Department, Southeast University Chengxian College, Nanjing 210088, China. Research Area: Recommender Systems.

