Summer 5-26-2017

Research on the Method of Friends Recommendation in Mobile Social Network Based on Multidimensional Similarity

Hongjiang Cao
School of Economics, Wuhan University of Technology, Wuhan, 430070, China

Li Xue
School of Economics, Wuhan University of Technology, Wuhan, 430070, China, xueliblackpink@126.com

Research on the Method of Friends Recommendation in Mobile Social Network Based on Multidimensional Similarity

Hongjiang Cao[1], *Li Xue*[2*]
School of Economics, Wuhan University of Technology, Wuhan, 430070, China

Abstract: Aiming at the issue of friend recommendation in social network, a friend recommendation method based on multi dimensional similarity is proposed. Compared with the existing methods, it is no longer limited to a single dimension of matching information, but from three dimensions of interest, time, and space to judge the similarity of users in different dimensionality, then it uses the similarity to do comprehensive evaluation to recommend friends to the target users. Experimental studies have shown that the precision rate of recommended results is close to 70 percent, and the checking efficiency rate is close to 60 percent when the method is applied to the recommendation service of mobile social network. The performance of this method is much higher than the friend recommended method based on single dimension and can be applied to a variety of characteristics of the mobile social network.

Keywords: mobile social network; friend recommendation; multidimensional similarity

1. INTRODUCTION

With the rapid development of mobile communication technology, mobile social network is also rapidly emerging in our lives. Mobile social network is a kind of location-based social network which is made up of mobile terminal and carrier. It makes the distance between virtual social and reality social closer and more relevant to the users. The association between users constitutes a user relationship diagram in which each user has his or her own habits and has some specific temporal and spatial characteristics. And the location association between users can form a location association graph. In the traditional social network, the two graphs are fragmented and unrelated, but the mobile social network can naturally get the user's social relations and location information at the same time, and can combine the virtual cyberspace and real physical space organically. Each user's space-time behavior and interaction are linked together through the entire mobile social network, truly reflecting the characteristics of the user in the real world. In the mobile social network, the mobile telephone network data has the natural data superiority, such as rich record properties, complete temporal and spatial information, frequent sampling, a more real social relationship, wide coverage and so on. Some of the representative social networking sites have become influential information platform. Users and information have exploded on social networks. As the popular and practical personalized service in social networking [1], friend recommendation's purpose is to recommend new friends to the user based on the user's existing friends and historical behavior, so that users (especially new users) can quickly establish a good friend relationship circle and assimilate into the social network information services, thereby increasing user activity and stickiness to social networks.

At present, friends recommended method is divided into two categories. One is based on the content of the friend recommendation [2]. This method predicts and searches for "friend" users with a high degree of matching according to the user's personal information and goal requirements. The other is based on collaborative filtering friend recommendation [3]. This method usually uses the nearest neighbor technique to estimate the degree of preference of the target user to the other user by the nearest user of the target user, so as to make friend recommendation. There are two kinds of problems in content-based recommendation method: 1) When the user

* Corresponding author. Email: xueliblackpink@126.com (Li Xue).

has very little registration information, incomplete information will directly affect the accuracy of the output results of part of friend recommendation method based on identity information. 2) It is difficult to guarantee the correctness of recommendation results by using only a single dimension of user characteristics (such as location information and user identity information).

In order to solve the above problems, this paper presents a content-based friend recommendation method. By analyzing the user 's behavior log, the method obtains the data of the user in three dimensions of time, space and interest, and judges the similarity degree of users in each dimension respectively. Finally, it can calculate a comprehensive index - "dissimilarity" between the various users as a basis for whether to recommend friends. Friend recommendation has been significantly improved in comprehensiveness and accuracy because of the use of the user's multidimensional information in this method. When calculating the "dissimilarity", adjusting the weight values of the three dimensions makes the method applicable to more types of social networks.

2. RELATED WORK

Friend recommendation in the social network recommend other users rather than the general merchandise, the recommended set of items is the user collection. Its recommended project collection is the collection of users. Recommended techniques in the traditional recommendation technology can be used as a friend recommended reference [4] [5]. But it cannot be simply applied to friend recommendation. Currently applications based on mobile social network can be divided into two categories: a class of applications designed to enhance the original interaction between friends. The other is to find new friends for the target user. In the application research to enhance the interaction between the original friends, [5] has studied the purely distributed friend relationship. Reference [6] considers the problem of supporting the users' contacts to communicate with each other in the functional design. In the research of finding new friends, [7] models the friendship model of mobile social networks based on location and uses the model to recommend new friends to the target users. Reference [8] analyzes the relationship between friendship network and the tendency of users through the analysis of Tencent, and proposes the network modeling based on user attributes. Reference [9] presents an approach of recommending friends by considering the real-time physical location proximity, offline behavior similarity and friendship network. In [10], it presents a simple but effective similarity-based prediction strategy based on label propagation. Reference [11] proposes a novel friend recommendation method ACR-FoF (algebraic connectivity regularized friends-of-friends) that considers both success rate and content spread in the network. In [12], it presents comprehensive empirical studies and verify the homophily of interest similarity across three interest domains and provides a practical prediction model under a real online social network environment.

In the method of content-based friend recommendation, [13] uses association rules to excavate the potential friends of the target users and make effective recommendation according to the user's age, region and income. In [14], a friend prediction model is established based on the location-related feature setting by analyzing the user's position list. Reference [15] proposes a social circle detection algorithm, which recommends new friends by defining the user's social circle similarity for the user. All of the above methods determine whether the friend relationship is established based on the user's characteristics similarity. Reference [16] proposes two private friend recommendation algorithms based on the social network structure and the users’ social tags.

In the friend recommendation based on collaborative filtering, [17] improves the users' preference predictions and the accuracy of the recommendation result by adding the socialization relationship of the potential users to the collaborative filtering algorithm. Based on the project attributes that the user has used, [18] further narrows down the search range of neighbor users by using collaborative filtering recommendation method based on user clustering. In addition, [13] uses the number and timing of meetings between users to calculate the degree of intimacy offline, which is used as an index of online friend recommendation.

3. THE METHOD OF FRIEND RECOMMENDATION FOR MOBILE SOCIAL NETWORK

3.1 Index definition and calculation method

Mobile Internet has mobility and immediacy, which is the biggest difference with the traditional Internet. Mobile users can use the mobile device to access the application at any time and any location. Location and time contexts are more closely related to mobile users. In the process of using of mobile applications, it results in a large amount of geographical information containing social media data.The spatial similarity between the two users is measured according to the positional information between the users. The interaction between friends is embodied in the real-time information communication. Therefore, this paper introduces the online time as the reference factor of friend recommendation, which is used to measure the similarity of the two users in time. Similar interests are two important factors that make two strangers become friends. Accurately grasping the interests of users can achieve accurate recommendation of friend. So this paper also uses the user's interest tags to measure the similarity of interest between two users. In order to measure the similarity of two users in three dimensions of space, time and interest, this paper proposes the following three indicators:

1) Interest similarity ($I_{i,j}$). The similarity of interest between any user i and j is expressed by $I_{i,j}$. Specific formula is as follows:

$$I_{i,j} = \frac{\sum_{i \in H, j \in N} P_{i,l} P_{j,l}}{\| P_i \| \times \| P_j \|} \tag{1}$$

In the formula (1): H represents the interest label sequence collection; N represents the user sequence collection; $P_{i,l}$ is the probability that the user has the interest l; $P_i = [P_{i,1}, P_{i,2}, \ldots, P_{i,k}]$ represents the probability vector of various interest tags owned by user i (k represents the total number of interest tags); $\|P_i\|$ is the norm of P_i. This paper prefers 2 – norm.

2) Geographic similarity ($L_{i,j}$). The spatial similarity of any user i and j is expressed by $L_{i,j}$, Specific formula is as follows:

$$L_{i,j} = \frac{\sum_{l \in P, j \in N} P_{i,l} P_{j,l}}{\| P_i \| \times \| P_j \|} \tag{2}$$

In the formula (2): P represents the set of base station sequences for communication; N represents a set of user sequences; $P_{i,l}$ represents the probability that user i will appear at location site l; $P_i = [P_{i,1}, P_{i,2}, \ldots, P_{i,m}]$ represents the probability vector that user i appears at each base station (m represents the total number of location sites). $\|P_i\|$ is the norm of P_i. This paper prefers 2 – norm.

3) Time similarity ($T_{i,j}$). It means the temporal similarity of users i and j. Specific formula is as follows:

$$T_{i,j} = \frac{\sum_{l \in T, i, j \in N} P_{i,l} P_{j,l}}{\| P_i \| \times \| P_j \|} \tag{3}$$

In the formula (3): T represents a set of time series sequences, and a natural day is divided into a plurality of time periods, all of which constitute a time series sequence set; N represents a set of user sequences; $P_{i,l}$ is the probability that user i is on-line in time period l; $P_i = [P_{i,1}, P_{i,2}, \ldots, P_{i,y}]$ represents the probability vector of user i on-line in each time period (y represents the total number of time periods). $\|P_i\|$ is the norm of P_i. This paper prefers 2 – norm.

In order to judge the similarity between any two users on the basis of the similarity of the above three users, this paper proposes a comprehensive index --- "user difference degree". Specific formula is as follows:

$$Differ_{i,j} = \sqrt{\omega_1 (I_{i,j} - \bar{I})^2 + \omega_2 (T_{i,j} - \bar{T})^2 + \omega_3 (L_{i,j} - \bar{L})^2} \tag{4}$$

In the formula (4), $Differ_{i,j}$ denotes the difference distance between user i and user j. ω1, ω2, ω3 represent the corresponding weights of the three similarity values and the range is [0, 1]. The weight can be adjusted according to the specific scene and the default is 1/3. I, L, and T respectively denote the mean value of the user's interest similarity $I_{i,j}$, geographic position similarity $L_{i,j}$ and time similarities $T_{i,j}$, Specific formula is as follows:

$$\bar{I} = \frac{n(n+1)}{2} \sum_{i \in N} \sum_{j \in N} I_{i,j} \tag{5}$$

$$\bar{T} = \frac{n(n+1)}{2} \sum_{i \in N} \sum_{j \in N} T_{i,j} \tag{6}$$

$$\bar{L} = \frac{n(n+1)}{2} \sum_{i \in N} \sum_{j \in N} L_{i,j} \tag{7}$$

In the formula (5), (6), (7): N represents the collection of user sequences. n represents the number of users.

3.2 User interest extraction and induction

Unlike geographical location information and online time information which can be extracted directly from the relevant log, user interest tags often need to be extracted from user-defined text with the widespread application of user-generated content in social networks. In order to more accurately calculate the similarity of interest among users, this paper constructs a two-level interest label classification system, and proposes a simple user interest label extraction method according to the classification system. Because of the commonality of user interest categories, this paper constructs a two-level classification system of interest tags in indirect way which estimates the interest of users by counting certain categories of public home pages and fan numbers on certain online social networking sites.

The specific content of constructing the classification system is to extract the public home page of the label, and delete fewer occurrences of labels according to the number of label statistics. The category labels for large categories are formed. And then non-significant sense of labels can be removed through artificial screening of the results. Finally, the classification system of the subclasses is obtained by merging similar labels. The results are shown in Table 1.

Table 1. Classification of interest tags

Category	Subclass	Category	Subclass
Food	Chinese food	Entertainment	Movie
	Fast food		Sport
	Beef		Music
	Desert		Travel
	Japanese Food		Sing
	Else		Else

The secondary interest label category above is used as primary keywords and sub-categories of interest as secondary keywords, and the user's interest tags are obtained by means of text retrieval matching. The acquisition process is shown in figure 1.

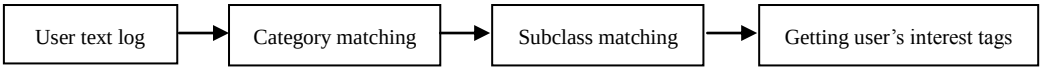

Figure 1. User interest label extraction

3.3 Steps for the implementation of the recommended method

Friend recommendation algorithm specific steps are as follows:

Step1: importing user text log, it calculates the interest similarity between any two users by extracting user interest tags.

Step 2: importing the user's online time log, it calculates the time similarity between any two users.

Step 3: importing the user's geographic log, it calculates the discrete geographical position similarity between any two users.

Step 4: it calculates the degree of difference between any two users according to the similarity of interest, time similarity and geographical position similarity.

Step 5: selecting the target user, it can recommend the other users ranked in the top -N (N values can be determined according to specific application scenarios) to the target user to become friend candidate according to the difference between the target user and other users distance.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4. 1 Experimental data and methods

In order to verify the effectiveness of this method, the real data set from Foursquare is selected in this experiment. The data set includes 501,940 anonymous user data in New York in February 2011 and 717,382 anonymous user data in Los Angeles, which contains basic information about the user, check-in information and friend relationship information. Since the data in the data set does not exist independently of the three parameters interest, time and location required in the proposed method of friend recommendation, three parameters need to be extracted before starting the experimental verification.

User check-in information format is shown in Figure 2. User ID represents the user ID. Tip indicates the specific contents of each user's attendance. The primary information fields include Street ID, Text, and Time. Street ID represents the street ID. Text is the user's text description of this check-in information. Time is the time of the check-in.

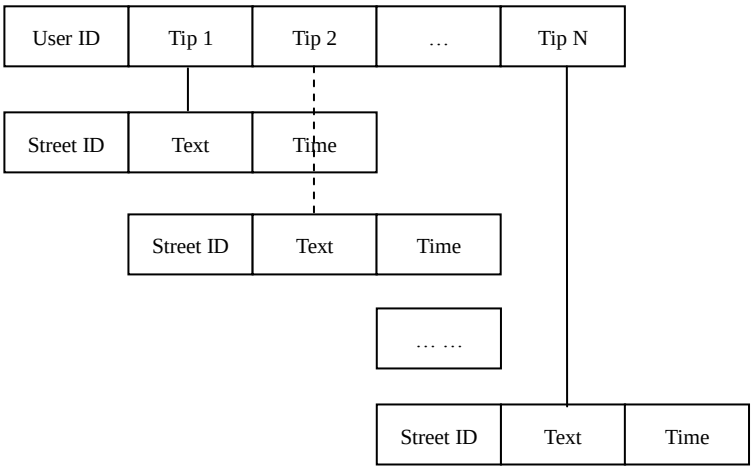

Figure 2. User check-in message format

This paper takes the Street ID as the location information and Time as the online time information to extract the user's interest label from the text Text. Taking the experimental data of NY and LA cities as an example, we compare recommended users by calculation with those of the target users. If the real friends included in the results are more, the effect of the recommended methods is better. Based on the concepts of precision and recall,

two evaluation indexes are proposed to evaluate the recommendation effect: P and R. P_{single} represents the precision of the proposed method for individual users. $P_{average}$ represents the average precision of the recommended method for all users. R denotes the Recall. $R_{average}$ represents the average recall of the recommended method for all users.

$$P_{\text{single}} = \frac{r}{f_{real}} \tag{8}$$

In the formula (8), R indicates the number of real friends of the target user included in the user to be recommended. f_{real} indicates the total number of real friends of the target user.

$$P_{average} = \frac{1}{N}\sum_{i=1}^{N}[P_{\text{single}}] \tag{9}$$

In the formula (9), N represents the total number of users.

$$R_{\text{single}} = \frac{r}{S} \tag{10}$$

In the formula (10), r indicates the number of real friends of the target user included in the user to be recommended. S indicates the total number of friends to be recommended.

$$R_{average} = \frac{1}{N}\sum_{i=1}^{N}[R_{\text{single}}] \tag{11}$$

In the formula (11), N represents the total number of users.

4. 2 Experimental results

The experiment treats other users whose degree of difference with the target user is the former one percent as recommended users, and the weight is set to the following four cases. case1(ω_1 = 1 /3, ω_2 = 1 /3, ω_3 = 1 /3); case2(ω_1 = 1, ω_2=0 /3, ω_3= 0); case3 (ω_1 = 0, ω_2 = 1, ω_3 = 0); case4(ω_1 = 0, ω_2 =0, ω_3 = 1). Among them, case1 represents a friend recommendation method based on multidimensional information, while the remaining three represent the friend recommendation method based on the single dimension information only in the user's interest information, the time information and the position information respectively. The results are shown in Table 2.

Table 2. Experimental data

City	Case	$P_{average}$	$R_{average}$
NY	case 1	0.8042	0.5799
	case 2	0.1814	0.1547
	case 3	0.1398	0.0832
	case 4	0.5186	0.4195
LA	case 1	0.7903	0.5472
	case 2	0.2015	0.1346
	case 3	0.1311	0.0912
	case 4	0.6316	0.4817

When the weight value of the three - dimensional parameter is ω_1 = 1/3, ω_2 = 1/3, ω_3 = 1/3, the average precision of the proposed result is close to 80%, and the checking efficiency is nearly 60%, which is obviously superior to other three. The main reason for this result is that Foursquare is a location-driven mobile social

platform. So compared to the friend recommendation only based on time or user interest, the recommended results of friend recommendation only based on the geographical position is more satisfying. But at the same time, we can see that the friends recommended results $P_{average}$ and $R_{average}$ increase nearly by 20% when online time and user interests are introduced based on the geographical location. Thus it can be seen that the friend recommendation results based on multidimensional data are significantly better than those of single dimension data.

In order to verify the effectiveness of the proposed method, we compare it with the recommendation method based on user's basic information and friend recommendation method based on mobile user social relationship mining. Specific experimental results are shown in Table 3.

Table 3. Comparison of recommendation effects of different methods

Method	$P_{average}$	$R_{average}$
Based on Multi - dimensional Similarity	0.7759	0.5717
Based on user profile information	0.2218	0.0377
Based on user socialization	0.5980	0.4135

According to the data of table 3, the friend recommendation method based on the multidimensional similarity is superior to the friend recommendation method based on the user's basic information and the friend recommendation method based on the social relationship mining. The main reason is that Foursqaure users only need to fill in the mailbox, gender and city of the three basic information when users perform user registration. The lack of basic information leads to poor recommendation results when the traditional method based on user's basic information to recommend friends is applied to the mobile social network. Compared with the recommendation method based on the user's basic information, the friend recommendation method based on social relationship mining has a great improvement in the recommendation precision and recommendation efficiency. But the method only refers to the social relationship between users, and the experimental data set is lack of further data to determine the intimacy of friends. So the recommendation result is still slightly worse than the friend recommendation method based on the multidimensional similarity.

5. CONCLUSIONS

This paper studies the friend recommendation in mobile social network. It recommends friends for target users from the interest, the time and the spatial dimensionality. The experimental results show that this method can get better recommendation results when applied to mobile social networks. Because the three-dimensional weight value can be flexibly adjusted according to the specific application scenarios, the method can also be applied to a variety of characteristics of the mobile social network. In the next work, time frequency and location frequency can be introduced into the calculation of similarity to further enhance the quality of recommended friends. In addition, only the interest label extraction and classification method based on the self-defined two-level interest label classification system is used in the experimental data processing. Therefore, in the future work, we can introduce the method of user interest mining in mobile social network. While extending the application scenario of the recommendation method, the reliability and accuracy of the recommendation results are further improved.

ACKNOWLEDGEMENT

This research was supported by the Hubei Provincial Natural Science Foundation Project(No.2013CFB350) and The Ministry of Education Humanities and Social Science Fund Project (No. 12YJC870002)

REFERENCES

[1] Liang Xiang. (2012). Recommendation System Practice. Beijing: Posts &Telecom Press. (in Chinese)

[2] B X WU, J XIAO. (2015). Potential friend recommendation based on user tagging. Journal of Computer Applications, 35(6): 1663-1667. (in Chinese)

[3] J HE, S L PAN, L HAN. (2015). Recommendation algorithm of SNS friends based on bilateral interest. Computer Engineering and Applications, 51(6): 108-113. (in Chinese)

[4] Z M ZENG, Xin Li. (2012). Personalized Information Recommendation Based on Context Awareness in Mobile Environment. Journal of Intelligence, 31(8): 166-170. (in Chinese)

[5] Jun Yang, J H Wu, Dan xiang Ai. (2013). A New Method of Multi - dimensional Information Recommendation Based on Scenario Similarity. Journal of The China Society for Scientific and Technical Information, 32(3): 262-269. (in Chinese)

[6] COUNTS S. (2007). Group-based mobile messaging in support of the social side of leisure. Computer Supported Cooperative Work, 16(2): 75-97.

[7] N LI, G L CHEN. (2009). Multi-layered friendship modeling for location based mobile social networks. Proceedings of the 2009 6th Annual International Mobile and Ubiquitous Systems. Piscataway, NJ: IEEE, 1-10.

[8] Wang Fulin, Gao Qiang, Liu Yanheng, & Wang Jian (2011). Modeling and Analysis of Buddy Network Topology in Instant Messaging System. Computer Science, 38 (8), 69-73. (in Chinese)

[9] Qiao, X., Jianchong, S. U., Zhang, J., Wangli, X. U., Budan, W. U., & Xue, S., et al. (2014). Recommending friends instantly in location-based mobile social networks. China Communications (English), 11(2), 109-127.

[10] Liu J, Xu B, Xu X, et al. (2016). A link prediction algorithm based on label propagation. Journal of Computational Science, 16, 43-50.

[11] Yu, Z., Wang, C., Bu, J., Wang, X., Wu, Y., & Chen, C. (2015). Friend recommendation with content spread enhancement in social networks. Information Sciences, 309, 102-118.

[12] Han, X., Wang, L., Crespi, N., Park, S., & Ángel Cuevas. (2015). Alike people, alike interests? inferring interest similarity in online social networks. Decision Support Systems, 69, 92-106.

[13] B GUO, Z YU, X ZHOU. (2012). Hybrid SN: interlinking opportunistic and online communities to augment information dissemination. Proceedings of the 2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing. Washington DC: IEEE Computer Society, 188-195.

[14] J CRANSHAW, E TOCH, J HONG. (2010). Bridging the gap between physical location and online social networks. Proceedings of the 12th ACM international conference on Ubiquitous computing. New York, 119-128.

[15] Y WANG, GAO L. (2014). Social circle-based algorithm for friend recommendation in online social networks. Chinese Journal of Computers, 37(4): 801-808. (in Chinese)

[16] Samanthula B K, Jiang W. (2015).Interest-driven private friend recommendation. Knowledge and Information Systems, 42(3):663-687.

[17] W H HUANG, X W MENG, L C WANG. (2011). A collaborative filtering algorithm based on users' social relationship mining in mobile communication network. Journal of Electronics and Information Technology, 33(12): 3002-3007. (in Chinese)

[18] G Y HUANG, LI Y C, GAO J P. (2010). Collaborative filtering recommendation algorithm based on user clustering of item attributes. Computer Engineering and Design, 31(5): 1038-104. (in Chinese)