# An Influence Calculating Method based on Mobile Phone Data

Yancui Shi[a], Yarui Chen[b]

College of Computer Science and Information Engineering, Tianjin University of Science & Technology, Tianjin, 300222, China

[a]email: syc@tust.edu.cn, [b]email:yrchen@tust.edu.cn

**Abstract.** In order to provide the personalized services, there is massive research about the mobile recommender system. The user influence plays an important role in the recommender system, so an influence calculation method based on mobile phone data is proposed in the paper. In the proposed method, we consider the user exchange behavior under context and the mobile user preference. Finally, the experimental results show that the proposed method can realize the influence calculation more accurately compared with the existing methods.

## 1. Introduction

Compared with the desktop, the smart phone has the following advantages: easy to carry, continuously improved function, low cost. Therefore, there are more and more users use the smart phone to communicate with other users or do shopping, entertainment, learning [1]. On the other hand, the smart phone has some defects, such as small screen, poor display ability, small memory and power of short duration. Hence, it is difficult for the mobile user to find the useful information from the information ocean.

In order to raise the mobile user satisfaction and increase the economic benefits of enterprises, the mobile recommender system is proposed according to the characteristic of the mobile network. But, due to the sparsity problem of the data, the accuracy of the recommender system is very low. In order to solve the problem, some research introduces the trust between users. However, there are few recommender systems that consider the user influence which plays an important role in recommender system. Therefore, it is necessary to realize the accurate calculation of user influence to improve the performance of the recommender system.

## 2. Related Work

The calculation methods of the user influence mainly are divided into three kinds: (1) based on the network topology; (2) based on user behavior; (3) based on the interaction information [2]. The calculation method based on network topology measures the influence mainly from two aspects: the nodes and the edges[3]. Whether the calculation is based on nodes or based on edge, they don't consider the user behavior, so the accuracy of obtained influence is relatively low. In order to improve the accuracy of obtained user influence, the researcher introduced the user behavior into the influence calculation, such as login behavior, generated information (e.g. review, forwarding information). Morales et al. calculated the user influence according to the number of forwarding information [4]. When constructing the model of the influence, Guo et al. regarded the log of user historical behavior in the social network as sample [5]. However, the above methods don't consider the exchange information between users, such as following, sharing or comment. In order to obtain the accurate influence, Anger et al. considered the communication content or message between users [6].

The above methods are mainly applied in Internet. In the light of the unique characteristic of mobile social network, an influence calculation method based on mobile phone data is proposed in the paper.

## 3. Proposed Method

In the section, the influence calculation method based on mobile phone data will be introduced in detail.

### A. *The influence based on user exchange behavior*

In the mobile network, the mobile user can communicate with the other users and use the mobile network services at anytime and anywhere. According to the mobile user exchange behavior, we can get more real social relationship than that in the Internet. In addition, the user exchange behavior will be impacted by the context. So we consider the user exchange behavior and the corresponding context information in the influence calculation.

The ways by which mobile user contacts with the other users contain the traditional ways (e.g., voice call, SMS) and the communication software (e.g., e-mail, QQ, WeChat, etc.). In the paper, we consider the duration and times of communication. The volume is calculated by the (1).

$$V_{u_i,u_j,c_k} = \sum_{l=1}^{N_w} \alpha_1 * \frac{L_{c_k,u_i,u_j,w_l}}{L_{c_k,u_j,w_l}} + \alpha_2 * \frac{T_{c_k,u_i,u_j,w_l}}{T_{c_k,u_j,w_l}} \tag{1}$$

In the type: $u_i \in U$ represents some specific mobile user, $U$ is the set of mobile users; $c_k \in C$ represents some specific context instance, $C$ represents the set of the context instances; $w_l$ represents some specific communication way, $N_w$ represents the types of used communication ways; $L_{c_k,u_i,u_j,w_l}$ represents the duration of the communication between the $u_i$ and $u_j$ under $c_k$ by the communication way $w_l$, $L_{c_k,u_j,w_l}$ represents the all duration of the communication of $u_j$ under $c_k$ by the communication way $w_l$; $T_{c_k,u_i,u_j,w_l}$ represents the times of the communication between the $u_i$ and $u_j$ under $c_k$ by the communication way $w_l$, $T_{c_k,u_j,w_l}$ represents the all times of the communication of $u_j$ under $c_k$ by the communication way $w_l$; $\alpha_1$, $\alpha_2$ represents the weight of the duration and times respectively, and $\alpha_1 + \alpha_2 = 1$.

According of the obtained volume, we can calculate the influence of $u_i$ toward $u_j$. The influence will increase with the increasing of the volume, while when the volume is bigger than a certain value, the growth rate of the influence slows down, and gradually tends to a fixed value. Hence, we choose the logarithmic function to measure the influence.

$$Ie_{u_i,u_j} = \sum_{l=1}^{N_c} \beta_{l_i} * \log_a^{V_{u_i,u_j,c_l}+1} \tag{2}$$

In the type: $a$ is the basic number of the logarithmic which is determined according to the range of volume; $N_c$ represents the number of the context instances; $\beta_{l_i}$ is employed to measure the weight of the context instance $c_l$ of $u_i$.

### B. *The influence based on user preference*

The influence of the user is not only related the user exchange behavior, and it is also related with the user preference. The target user will influence the preference of the other users that have similar preferences with target user. the more the preference of the user is similar with the target user, the greater the influence of the target user is. So, we employ the similarity between two users to measure the influence of the target user. In addition, we consider the timing problem of the user preference in the paper. The reason is that the user only influence the preference of the users who use the mobile network service later than the target user.

In order to obtain the user similarity, we need to learn the user preference according to the user behavior. Firstly, we compute the usage volume of the mobile user by (3).

$$V_{u_i,s_j,c_k} = \lambda_1 * \frac{L_{u_i,s_j,c_k}}{L_{c_k,u_i}} + \lambda_2 * \frac{T_{u_i,s_j,c_k}}{T_{c_k,u_i}}$$

(3)

In the type: $s_j \in S$ represents the mobile network service $j$, $S$ is the set of the mobile network services; $L_{u_i,s_j,c_k}$ represents the duration that $u_i$ uses $s_j$ under $c_k$, $L_{c_k,u_i}$ represents the duration that $u_i$ uses the all mobile network services under $c_k$; $T_{u_i,s_j,c_k}$ represents the times that $u_i$ uses $s_j$ under $c_k$, $T_{c_k,u_i}$ represents the times that $u_i$ uses the all mobile network services under $c_k$; $\lambda_1$, $\lambda_2$ are parameters and are employed to measure the wight of the duration and times respectively, and $\lambda_1 + \lambda_2 = 1$. And then we use the logarithmic function to calculate the user preference.

$$P_{u_i,s_j,c_k} = \log_{a_1}^{V_{u_i,s_j,c_k}+1} \tag{4}$$

In the type: $a_1$ is the basic number of the logarithmic function.

Based on the obtained the user preference, we employed the improved Pearson Correlation Coefficient to calculate the influence. The formula is represented as follows:

$$Ip_{u_i,u_j} = \frac{\theta * \sum\limits_{c_l \in C} \sum\limits_{s_k \in S_{tl}} \beta_{l_i} * \beta_{l_j} (P_{u_i,s_k,c_l} - \overline{P_{u_i,c_l}})(P_{u_j,s_k,c_l} - \overline{P_{u_j,c_l}})}{\sqrt{\beta_{l_i} * \sum\limits_{s_k \in S_{tl}} (P_{u_i,s_k,c_l} - \overline{P_{u_i,c_l}})^2 \sum\limits_{s_k \in S_{tl}} \beta_{l_j} * (P_{u_j,s_k,c_l} - \overline{P_{u_j,c_l}})^2}} \tag{5}$$

In the type: $S_{tl}$ represents the set of common mobile network services that $u_i$ and $u_j$ both used under $c_l$ and $u_i$ used these mobile network services sooner than the $u_j$; $\overline{P_{u_i,c_l}}$ is the mean value of the user preference under $c_l$; $\theta$ is employed to measure the effect of the number of the common mobile network services, it can be obtained by (6).

$$\theta = \frac{|S_t|}{|S_{u_j}|} \tag{6}$$

In the type: $S_{u_j}$ represents the set of the mobile network services that $u_j$ used; $S_t = \{s_k | s_k \in S_{u_i}$, and $s_k \in S_{u_j}, t_{u_i,s_k} < t_{u_j,s_k}\}$, $t_{u_i,s_k}$ represent the time that $u_i$ used $s_k$.

## C. The influence between users

In the section, we calculate the influence between users according to the above works.

$$I(u_i, u_j) = \mu_1 * Ie_{u_i,u_j} + \mu_2 * Ip_{u_i,u_j} \tag{7}$$

In the type: $\mu_1$, $\mu_2$ are weight, and $\mu_1 + \mu_2 = 1$.

## 4. Experiments and Analysis

However, there is no influence information in the dataset. So, we can't verify the proposed method directly. The influence calculation method is proposed mainly to apply into recommender system in the paper. Hence, we introduce the influence into the recommender system that is based on the traditional collaborative filtering.

## A. Dataset

We use the real dataset to verify the validity of the proposed method in the paper [7]. The dataset includes the behavior information of 94 mobile users and 200 kinds of mobile network services in nine months.

## B. Evaluation

The evaluation methods include $MAE_p$ and $MAE_r$. The $MAE_p$ is as follows.

$$MAE_p = \frac{\sum\limits_{c_k \in C'_{u_i}} \sum\limits_{s_j \in S'_{u_i}} |P_{u_i,s_j,c_k} - P'_{u_i,s_j,c_k}|}{N_p} \tag{8}$$

In the type: $N_p$ represents the number of the predicted preferences; $S'_{u_i}$ is the set of mobile network services that the target user don't use but his nearest neighbors have used under $c_k$; $C'_{u_i}$ presents the set of context instances under which the target user don't use the mobile services but his nearest neighbors have used.

The $MAE_r$ is as follows.

$$MAE_r = \frac{\sum\limits_{h=0}^{N_0} |r_i - r'_i|}{N_0} * \frac{N_{top}}{N_0} \tag{9}$$

In the type: $N_{top}$ represents the number of recommended the mobile network services; $N_0$ represents the number of the mobile network services that are in the list of the mobile network services that the target user used; $r_i$ is the rank of the recommended mobile network services; $r'_i$ is the rank of the mobile network services that the mobile user used.

### C. Experimental step

We choose the traditional collaborative filtering (CF) as the benchmark method. The number of the mobile network services that we recommend is 10, that is $N_{top}=10$.

Step 1. Introduce the influence of user exchange behavior (method 1). Where, $\alpha_2$ is set 0.2,0.3,0.4,0.5,0.6,0.7,0.8; the value of the influence is set in [0,1], and the setting of $a$ is limited by the communication volume, that is $V_{u_i,u_j,c_k}$, so $a$ is set (the maximum of the telecommunication volume+1); $\beta_{l_i}$ is set according to the results in [8].

Step 2. Introduce the influence of user preference (method 2). $\lambda_1$ is set according to the results Step 1. directly. Because the value of $V_{u_i,s_j,c_k}$ is in [0,1] and the value of the preference is positive, so we set $a_1=2$; $\mu_1$ is set 0.2,0.3,0.4,0.5,0.6,0.7,0.8.

### D. Experimental results and analysis

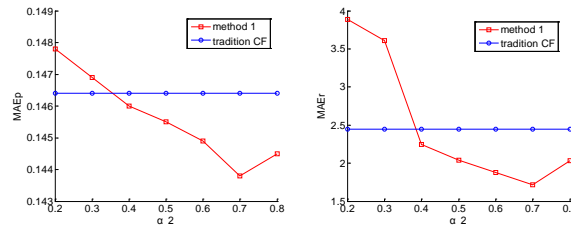*1) the impact of the parameter $\alpha_2$*



Fig.1 The results obtained by the traditional CF and method 1 when $\alpha_2$ is set different values

From Fig1., we can know that: (1) compared with the traditional CF, when $\alpha_2 > 0.3$, the method 1 can obtain the better results. This is because when $\alpha_2 < 0.3$, the weight of the exchange times is very small, so leads the obtained influence is not accurate. When $\alpha_2 > 0.3$, the obtained influence is improved, so the recommendation results become better. (2) when $\alpha_2 = 0.7$, the obtained result by method 1 is the best. The result shows that the exchange times plays a more important role than the exchange duration.

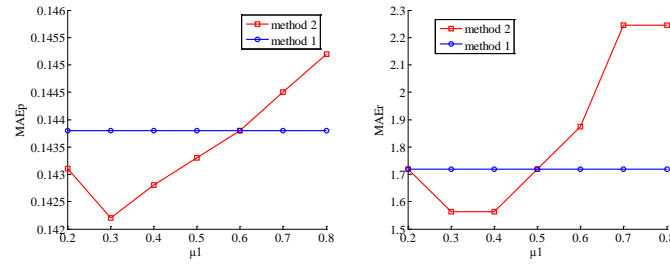*2) The impact of the weight $\mu_1$ and $\mu_2$*

Fig.2 The results learned by the method 1 and method 2 when $\mu_1$ is set different values

The Fig.2 shows that when $\mu_1$=0.3, the result is the best. According to the results, we can get the conclusion, the the user preference is more important than the user exchange behavior.

## 5. Conclusion

In the mobile network, it is feasible to collect the user exchange behavior and learn user preference. Hence, in order to improve the accuracy of the influence, the user exchange behavior and the user preference are considered in the paper. The experimental results show the proposed method is better than the existing methods.

## Acknowledgement

## References

[1] Zhipeng Liu, Dechang Pi. Mining social influence of nodes from mobile datasets [J]. Journal of Computer Research and Development, 2013 50 244-248.

[2] Xindong Wu, Yi Li, Lei Li. Influence analysis of online social networks [J]. Chinese Journal of Computer, 2014 37 (4) 735-752.

[3] Jimeng Sun, Jie Tang. A survey of models and algorithms for social influence analysis [C]. Social network data analytics. Springer US, 2011 177-214.

[4] A.J. Morales, J. Borondo, J.C. Losada, et al. Efficiency of human activity on information spreading on Twitter [J]. Social Networks, 2014 39 1-11.

[5] Jing Guo, Yanan Cao, Chuan Zhou, et al. Influence weights learning under linear threshold model in social networks [J]. Journal of Electronics & Information Technology, 2014 36 (8) 1084-1089.

[6] Isabel Anger, Christian Kittl. Measuring influence on Twitter [C]. In: Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies. ACM, 2011.

[7] Nathan Eagle, Alex Pentland, David Lazer. Inferring social network structure using mobile phone data [J]. the National Academy of Sciences (PNAS), 2009 106 (36) 15274-15278.

[8] Yancui.Shi, Xiangwu Meng, Yujie Zhang, et al. A trust calculating algorithm based on mobile phone data [C]. In: Proceedings of the IEEE Global Telecommunications Conference, (GLOBECOM). 2012 2012-2017.