Behavior Prediction of Mobile User Based on Context Similarity

Chengfang Tan, Qixiang Song and Xiaoyin Wu

***Abstract*—**Contextual information of users in mobile environment is diverse. Contextual factors have different importance to users in choosing resources. This paper presents a context-based mobile user behavior analysis model. Firstly, by using the context information entropy to calculate the importance of different context attributes, context similarity can be calculated. Then, combining with the traditional collaborative filtering recommendation algorithm, the target user's nearest neighbor set is obtained based on user similarity. Finally, the behavior preference of the target user can be predicted, and a target user's recommended list of projects also generate under the current context environment. Experimental study is carried out and the results show that the proposed prediction method have a better performance in a specific context.

***Keywords*—**context, mobile user, user behavior, personalized service.

I. INTRODUCTION

Context-aware services are generated in such a context and affect people's lives. The information obtained by the mobile terminal sensor can be used to identify the current context, and finally make a more accurate prediction of the user's behavior through certain analysis and reasoning. Context aware need to take account of the changes of users' needs and external environment in time dimension, so as to provide users with personalized services. Therefore, how to effectively obtain contextual information and forecast the behavior of users has become a hot topic in the field of personalized service.

Many research achievements have been achieved in the field of context awareness and mobile user behavior analysis at home and abroad. Among them, context awareness is widely used in the commercial wide [1]. Springer et al. proposed logistics information based on context awareness, using context aware technology to provide more accurate logistics information for users [2]. Helin proposed a Context Awareness based Service Coordination model (CASCOM), which provides users with remote monitoring and context-aware applications such as e-convergence, shopping and emergency medical services through intelligent agents [3]. Zaslavsky put forward the application scenarios and problems of context awareness technology in mobile assistant business, including contextual discovery, extraction, understanding, dissemination [4]. Yamakami proposed a method based on the mobile terminal user log to analyze the user's behavior [5].

From the current research on user behavior analysis, the research on context-based user behavior prediction is still at a preliminary stage. There are some problems such as the lack of quantification of context weight and the lack of accuracy of user behavior prediction. Based on the analysis of context information, this paper predicts the items that the target user is interested in according to the similar neighbor selection behavior of the item under similar context. Using the current context as a constraint, the set of neighbor users that is most similar to the target user in the current context or the similar context is identified. Thereby providing the selected resources of the neighboring users for the target users, the user can obtain better experience and perception of context through the mobile terminal.

II. Theory of context

Context refers to any information that can describe the characteristics of entities. The so-called entity refers to human, location, physical or virtual objects that interact with users and applications, including users and its application. B.N.Schilit proposed context awareness and define the word context as three types: user context, physical context and computing context [6].

On the basis of mobile computing, the mobile terminal is regarded as the center of geographic location or network topology. By making full use of the environmental information, context aware computing takes the information in the environment of the terminal as the content of computation, and tries to make the service discovery and service implementation process finished by the terminal itself [7]. Therefore, the content and the way of service are usually closely related to the context information, and also change with the change of users' context, so that we can realize context aware service. Context aware computing uses sensors and related technologies to enable application devices to perceive the current context, thereby allowing applications to better adapt to context.

III. User preference analysis based context

1. Calculation of context information entropy

Because information entropy can measure the uncertainty of random variables in the system, this research explores how to

This work was supported by Key Projects of Excellent Young Talent Support Program in Colleges and Universities of Anhui Province (No.gxyqZD2016344) and University Students Innovative Undertaking Training Program of Anhui Province (No.201710379021) and Teaching Research Project of Suzhou University (No.szxy2016jy32).

Chengfang Tan is with School of Information Engineering, Suzhou University, Suzhou 234000, Anhui, China (corresponding author; e-mail: 874036730@qq.com).

Qixiang Song is with School of Information Engineering, Suzhou University, Suzhou 234000, Anhui, China.

Xiaoyin Wu is with School of Information Engineering, Suzhou University, Suzhou 234000, Anhui, China.

identify important context factors that affect user's selection behavior or preference. By calculating the importance of these context factors on users, we further analyze user preferences based on these important context factors.

Under a specific context instance c_{kq}, user selects the entropy value of the item whose attribute category is a_i . User's selection of the item for a certain category under this context instance is expressed. The specific formula is as follows:

$$E(c_{kq}) = \sum_{i=1}^{n} p_{c_{kq}}(a_{ij}) \log_n p_{c_{kq}}(a_{ij}) \tag{1}$$

Where $E(c_{kq})$ represents the item selection condition of the user for different attribute categories under the context instance c_{kq} . And $p_{c_{kq}}(a_{ij})$ represents the probability of items belonging to the certain feature a_{ij} among the items of the attribute category a_i, which selected by the user u under the context instance c_{kq}. And n is the number of attribute categories that the project has.

Due to the different distribution of context samples contained in context factors, the conditional entropy is used to calculate the context information entropy corresponding to the context factor c in this paper. That is, the entropy of corresponding context factors c_k under different context instances distribution is calculated. It is as shown in the following formula:

$$I(c_k) = \sum_{q=1}^{t} p(c_{kq}) E(c_{kq}) \tag{2}$$

Where $I(c_k)$ represents the degree of disorder of user's choice for the project under the distribution of the different context instance c_{kq}. $p(c_{kq})$ is the distribution value of the c_k sample under the context instance c_{kq}. t is the number of the context instance sample contained in the context factor.

Since the size of the entropy of the contextual factor c_k is determined by the entropy and its distribution of all instances it contains, the range of $I(c_k)$ is between 0 and 1 according to the range of $E(c_{kq})$.

2. User preference calculation based on context

By calculating the context information entropy, it is possible to identify the importance of context factors to users, and then identify the context factors that can have an important impact on user preferences.

This paper introduces the concept of context importance to describe the importance of context factors for users, and analyzes user preferences based on the importance of the context factors and the context instances it contains. The calculation formula of context importance is as follows:

$$H(c_k) = 1 - I(c_k) \tag{3}$$

Where c_k is the context factor that has a significant impact on user preference. $H(c_k)$ is the importance of the context factor.

Therefore, the preference of user u for item attribute a_{ij} under a set of context information CG is calculated as follows:

$$P_u(C, a_{ij}) = \sum_{k=1}^{x} p(a_{ij} \mid c_{kq}) \times H(c_{kq}) \tag{4}$$

Where c_{kq} is an instance that can have a significant influence on the preference of user in the group of context information CG. $p(a_{ij} \mid c_{kq})$ is the preliminary preference value of user for item with feature a_{ij} under the category of attributes a_i. x is the number of context instances that have a significant impact on user preferences.

IV. Behavior analysis of mobile user based on context

1. Data model representation

The algorithm in this paper is based on U-I-C three-dimensional data model. In this model, users, items and contexts represent one dimension respectively, which is represented by vector space composed of their respective attributes. Each attribute has its own specific attribute values.

2. Calculation of user similarity

The similarity between the target user u and any user v is defined as $as(u,v)$ for the similarity of the item attribute preference. according to the method of calculating user preferences above, we get the preference degree of users u and v for the different feature items of the attribute category a_i under the context information C, which expressed as $P_u(C, a_i)$ and $P_v(C, a_i)$ respectively. The cosine similarity is used to calculate the similarity between the target user u and the user v [8]. The calculation formula is as follows:

$$as(u,v) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{h} p_u(C, a_{ij}) \cdot p_v(C, a_{ij})}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{h} p_u(C, a_{ij})^2 \cdot \sum_{i=1}^{n}\sum_{j=1}^{h} p_v(C, a_{ij})^2}} \tag{5}$$

Where n is the number of attribute category owned by the project. And h is the number of features contained in the i-th attribute. a_{ij} is the j-th feature value under the i-th attribute category.

In addition, this study collects the common scoring items from different users as the nearest neighbor. Assuming that the item set $I_{uv} = I_u \cap I_v$ jointly evaluated by the user u and the user v, the Pearson correlation between the user u and the user v can be defined as:

$$ps(u,v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \overline{r_u})^2 \sum_{i \in I_{uv}} (r_{v,i} - \overline{r_v})^2}} \tag{6}$$

Where $r_{u,i}$ represents the score of user u for the item i. Similarly, $r_{v,i}$ represents the score of user v for the item i. $\overline{r_u}$

and $\overline{r_v}$ respectively indicates the average score of user u and user v to the item i .

This research considers the similarity between users from two part of user scoring and attribute preference. Therefore, using the linear weighted function, the similarity between user u and v is calculated as follows:

$$sim(u,v) = \lambda as(u,v) + (1-\lambda) ps(u,v) \tag{7}$$

Where λ is the adjustable coefficient of [0, 1]. When the value is 1, the similarity degree between users is calculated from the perspective of users' preferences for item attributes. When the value is 0, the similarity between users is calculated by user's scoring data.

3. Calculation of context similarity

Generally, the user's evaluation of the project resources is generated in a group of related contexts. And there are few scoring records in a single dimension context instance. So, the following formula is used to calculate the preference score of the nearest users.

$$s_{v,I,C} = \sum_{i=1}^{n} p(a_{ij} \mid c_{kq}) \cdot w_i \tag{8}$$

Where a_{ij} is the attribute feature of the item *i*. $p(a_{ij} \mid c_{kq})$ is the preference degree of user *v* for the attribute feature under the single dimension context instance. w_i is the weight of the i-th attribute feature.

According to each important context factor, we measure the similarity of two different context instances based on the score values of the common scoring items. The formula is allows:

$$sim_{c_k}(c,c_x) = \frac{\sum_{I \in I_x} (r_{v,I,c} - \overline{r_{v,c}})(r_{v,I,c_x} - \overline{r_{v,c_x}})}{\sqrt{\sum_{i \in I_x} (r_{v,I,c} - \overline{r_{v,c}})^2 \sum_{i \in I_x} (r_{v,I,c_x} - \overline{r_{v,c_x}})^2}} \tag{9}$$

This formula expresses the similarity of two different context instances c and c_x for user *u* under the context factor c_k . Where I_x is the common scoring item collection of the neighbor user *v* under two different context instance $\overline{r_{v,c}}$ and $\overline{r_{v,c_x}}$ respectively indicate the average score of the common scoring project under the context instance c and c_x .

4. Prediction of user behavior

Based on the idea of user collaborative filtering recommendation, this paper introduces the context similarity to collaborative filtering algorithm. Under the context of single dimension context instance *c*, the prediction score of target user *u* for the item *i* is mainly calculated according to the score of *i* for its neighbor users *v* in similar context instances. Therefore, the corresponding prediction score function is as follows:

$$r_{u,I,c} = \frac{\sum_{v \in N_u, c_x \in N_c} r_{v,I,c_x} \times sim(u,v) \times sim_{c_k}(c,c_x)}{\sum_{v \in N_u, c_x \in N_c} | sim(u,v) \times sim_{c_k}(c,c_x) |} \tag{10}$$

Where $c_x \in N_c$ is the neighbor set that most similar to the current context instance *c*. $v \in N_u$ is the neighbor set of the target user *u*. r_{v,I,c_x} is the score of the item i of the neighbor user *v* under similar context instance c_x .

Context weight of each context instance *c* is $H(c_k)$ under the context information *CG*. The scoring of the item is calculated under single dimension context instance. The prediction formula is as follows:

$$R_{u,I,CG} = \sum_{c \in CG, c_k \in CG} r_{u,I,c} \times H(c_k) \tag{11}$$

Where $r_{u,I,c}$ is the predictive value of the target user u on the item i under the single dimension context instance *c*.

V. Experiments and Analysis

1. Experimental data

This paper uses the real dataset of local famous catering consumption to test the performance of the algorithm. The dataset contains basic user context information, including gender, age, geographic location, dining preferences, user location records, and historical consumption records and so on. In order to verify the result conveniently, we extract some context data from the dataset, which contain 2974 ratings of 157 users for 139 restaurants. The data is divided into a training set and a test set according to the ratio of 3: 1.

2. Evaluation index

In the experiment, the widely used Mean Absolute Error (MAE) is used as the evaluation criterion for the accuracy of the recommended algorithm [9]. MAE is used to calculate the accuracy of the algorithm by calculating the deviation between the predicted user score and the actual user score. The smaller the MAE value, the higher the quality of the algorithm [10].

The definition of user scoring prediction set is expressed as $\{p_1, p_2, \ldots, p_N\}$. The corresponding actual user score corresponding is expressed as $\{r_1, r_2, \ldots, r_N\}$ .The MAE is defined as follows:

$$MAE(N) = \frac{\sum_{i=1}^{N} |p_i - r_i|}{N} \tag{12}$$

3. Analysis of experimental results

In this paper, λ is used to adjust user similarity $as(u,v)$ based on attribute preference and user similarity $ps(u,v)$ based on item scoring. If the value of the weight coefficient λ is too large or too small, it will affect the recommendation accuracy. So it is necessary to examine the influence of different weight coefficients on the recommended accuracy. Because of $\lambda \in [0,1]$, λ is measured from 0 to 1 in the experiment. The relationship between MAE and the value of λ is as shown in Figure 1.

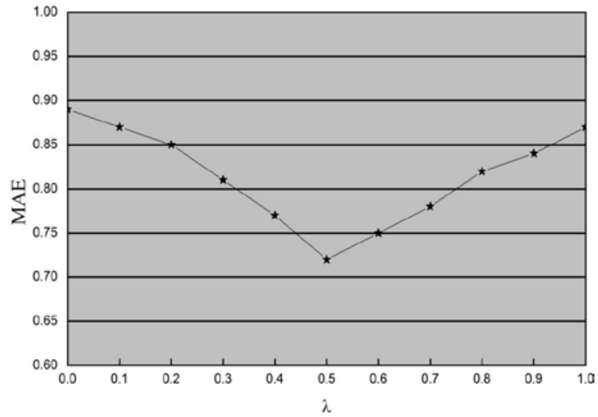

Fig. 1. The corresponding MAE value under different λ

As can be seen from the Figure 1, when the value of the weight coefficient is 0.5, the MAE value is the smallest. And the recommendation system has the best quality at this time. So we set $\lambda = 0.6$ in subsequent experiments.

In order to verify the validity of the proposed model, we select the context factors that have an important impact on user preferences. The following two methods are compared: (1) The proposed collaborative filtering recommendation based on user context (UC-CF). (2) The traditional collaborative filtering recommendation based on user (User-CF).

In the actual context of user behavior prediction, the problem of data sparsity is often encountered. So the accuracy of the algorithm is affected. In this experiment, the different sparsity and the number of different neighbors are selected to compare the experiments. This study chose on different numbers of neighbors in experiment. The number of neighbors gradually increases from 5 to 30 at 5 intervals. The MAE values obtained in the experiment are shown in the following table:

Table 1. Comparison of two different algorithms

nn	Density=0.1		Density=0.15		Density=0.2	
	UC-CF	User-CF	UC-CF	User-CF	UC-CF	User-CF
5	0.88	0.95	0.87	0.92	0.84	0.9
10	0.83	0.89	0.81	0.85	0.78	0.83
15	0.68	0.72	0.63	0.68	0.6	0.65
20	0.72	0.73	0.68	0.71	0.64	0.69
25	0.76	0.81	0.74	0.78	0.73	0.72
30	0.78	0.84	0.76	0.79	0.75	0.76

When the data sparsity is 0.1, the comparison between the proposed algorithm and the traditional user-based collaborative filtering is shown in Figure 2.

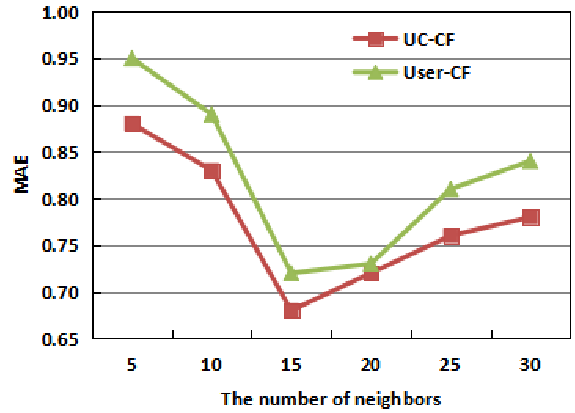

Fig. 2. MAE comparison between UC-CF and User-CF

It can be seen from the Figure 2 that the average absolute error of the prediction results obtained by the two methods will decrease as the number of neighbor users increases. Therefore, as the number of neighbor users gradually increases, the prediction accuracy of two recommended methods will be higher and higher. However, when the number of users reaches a certain number, the values obtained by the two methods show an increasing trend and the prediction accuracy of the recommended decrease gradually. Moreover, when the number of the neighbor of the target user reaches 20, the predicted values of the two methods are all the lowest.

VI. Conclusions

In this paper, the problem of user behavior prediction in mobile environment is studied. Under the mobile Internet environment, the user's demand preference is closely related to the specific context in which the user is located. The research work in this paper mainly focuses on the acquisition and processing of contextual information of mobile terminal user and user behavior prediction. The experimental results show that this method can better obtain the user's behavior preference, and help to improve the prediction effect of the personalized recommendation system. In addition, contextual information has very high complexity and diversity. In order to facilitate the calculation and improve the prediction efficiency, we only draw a few relatively influential context factors for experiment and analysis, which resulted in the lack of information in a certain extent, affecting the accuracy of calculation results. Therefore, in the future research, we can conduct further research on pretreatment of the high dimensionality context information and dimension reduction.

References

[1] Chiu P H, Kao G Y M, Lo C C, “Personalized blog content recommender system for mobile phone users”, *International Journal of Human-Computer Studies*, vol. 68, no.8, pp. 496-507, 2010.

[2] Springer T, Kadner K, Steuer F, et al. “Middleware support for context-awareness in 4g environments”, *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*, pp. 203-211, 2006.

[3] Helin H, Klusch M, Lopes A, et al. “Context-aware business application service co-ordination in mobile computing environments”, *Proceedings of the International Workshop on Ambient Intelligence*, pp. 1-12, 2005.

[4] Zaslavsky A, "Mobile agents: can they assist with context awareness", *2013 IEEE 14th International Conference on Mobile Data Managemen*t, IEEE Computer Society, pp.304-304, 2004.
[5] Yamakami T, "Unique identifier tracking analysis: A methodology to capture wireless internet user behaviors", *Proceedings. 15th International Conference on. IEEE*, pp.743-748, 2001.
[6] Kim J., Lee D., Chung K.Y, " Item recommendation based on context-ware model for personalized u-healthcare service", *Multimedia Tools and Applications*, vol. 71, no.7, pp. 855-872, 2014 .
[7] Colombo L. O., Valencia G.R. Rodriguez G. A., et al. "RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes", *Expert Systems with Applications*, vol. 42, no.3, pp.1202-1222. 2015..
[8] Yogesh S.R., Mohan S.K, "ClickSmart: A Context-Aware Viewpoint Recommendation System for Mobile Photography", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no.1, pp.149-158, 2017..
[9] Liu Q, "Accurate and Diverse Recommendations Based on Communities of Interest and Trustable Neighbors", *International Journal of Security and Its Applications*, vol. 9, no.3, pp. 63-76, 2015.
[10] Maria C. R., Sergio I, "Pull-based recommendations in mobile environments", *Computer Standards & Interfaces*, vol. 44, no.1, pp. 185-204, 2016.

Chengfang Tan was born on Feb. 16, 1981**.** She received the Master degree in computer science and technology from Nanjing Normal University of China. Currently, she is a researcher (assistant professor) at Suzhou University, China. Her major research interests include information processing and data mining. She has published many papers in related fields.

Qixiang Song was born on Oct. 12, 1964**.** He received the Master degree in electronic technology from Anhui University of China. Now, he is a professor at Suzhou University, China. His major research interests include information security and image processing.

Xiaoyin Wu was born on Sept. 11, 1979**.** She received the Master degree in computer science and technology from Hefei Polytechnic University, China. . Her major research interests include artificial intelligence. She has published many papers in related fields.