Analysis of Mobile Internet Multi - context User Preference

YIN Zhenyu[1, a],Jiang Yi[1, b], He Jun[2, c]

[1]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[2]Chongqing ZTE Research Institute, Chongqing 401121, China

[a]1461816223@qq.com, [b]jiangyi@cqupt.edu.cn, [c]he.jun@zte.com.cn

Keywords: mobile Internet,amorous environment,user behavior,association rules,preference model.

Abstract: With the increasing popularity of smart mobile devices and the rapid development of mobile Internet, which provides mobile users with a broader platform for the behavior. The rich data connotation provides favorable conditions for more accurate user behavior analysis. Based on the behavior of mobile Internet users, this paper proposes a multi - context mobile user preference analysis method. This method classifies the user preference content according to the user information and the log to improve the efficiency of data analysis. And the user preference correlation model is generated by using the association rule Apriori algorithm in SPSS Modeler to construct the user preference analysis model. Finally, the effectiveness of the method is verified by analyzing the examples and comparing experiments.

1. Introduction

With the development of the information society, intelligent mobile devices are becoming more and more popular. Mobile Internet users will have different access behavior in different contexts.

Compared with the traditional Internet users,The situation will be more diversified, The user preference analysis in multiple contexts will promote the development of personalized recommendation business, At the same time it has more and more commercial value.In order to improve the accuracy of mobile user preferences, Tveit et al. Analyzes the behavior of mobile users by contextual user preference extraction and context perception [1]. Hu Muhai et al. Proposed a general research framework based on context-aware user preference analysis, defined the contextual meaning in detail, and studied the user preference model in the contextual data model and context [2]. Meng Xiangwu and other research on the demand of mobile users in the context of classification, given the basic framework of mobile user needs analysis technology [3] .The YAO through the Apriori algorithm to tap the mobile Internet value-added services in order to obtain mobile users to mobile value-added business preferences [4].

The studies on the needs of mobile Internet users above have achieved good results, but most do not consider the complexity and diversity of mobile user scenarios fully [5]. In addition, in the mobile Internet user preference analysis, user context perception, user preference classification measurement, analysis and verification and application services is also worthy of further study and discussion [6]. This paper presents an analysis method of interest degree combined with user context information. this method can effectively measure the user's preference and tap the user preferences in different situations.

2. Mobile user preference analysis

Compared to the traditional Internet, mobile Internet users access Internet time is more fragmented and browsing behavior is random, so it’s more difficult to dig to useful information; The boundaries of mobile users' lives and work are becoming increasingly blurred, As the user moves, Its geographical location changes will lead to user needs tend to diversity [7]. When a user visits a Web site, the user categorizes the content of the user's Web site Combined with the user's access to

information and situational factors to Analyze the content and preferences of the user at different times and locations.

2.1 Related definitions

Based on the above analysis, this paper gives the Related definitions of User preference for interest and context.

Definition 1 location and time are two key factors that affect the user's interest In a mobile Internet environmen. The context information of this paper only considers the two factors of location and time [8], User i's context information is defined as (1)

$$situation(i) = \{location(i), time(i)\} \tag{1}$$

Where $location(i)$ is the current location information of the user, indicating the time of the current user's visit to the Web site. In this paper, the user's context information is shown in Table 1.

Table 1 Example of the user's context information

Serial number	Web site	time	location
1	c_1	morning	l_1
2	c_2	afternoon	l_2
3	c_3	evening	l_3
4	c_4	early morning	l_2

Where l_1, l_2, and l_3 represent different geographical locations. Time information is divided into the morning (6: 00-12: 00), afternoon (14: 00-18: 00), evening (19: 00-24: 00), early morning (24: 00-5: 00).

Definition 2 The similarity between the current situation of user i and the user's historical context can be defined as (2)

$$Sim(situation_c(i), situation_h(i)) = \frac{num(s_1, s_2, ..., s_k, ..., s_n)}{n \times N} \tag{2}$$

Where s_k represents a single context of user i, $s_1, s_2, ..., s_k, ..., s_n$ represents multiple contexts, such as time context, geographic location, weather context, etc., and $num(s_1, s_2, ..., s_k, ..., s_n)$ represents the sum of the number of occurrences of $s_1, s_2, ..., s_k, ..., s_n$ in user i's historical context. n represents the number of context attributes, and N represents the number of times the user has visited [9].

Definition 3 The degree of interest（c_j）of the mobile user i to an access content（$P_i^{c_j}$）is defined as (3)

$$P_i^{c_j} = \omega_1 \times F_i^{c_j} + \omega_2 \times T_i^{c_j} \tag{3}$$

Where $F_i^{c_j}$ and $T_i^{c_j}$ are two dimensions that calculate the degree of interest that the user has access to the content. ω_1 and ω_2 are the weights of these two dimensions [10], and the weights are in the range of 0 to 1 and satisfy $\omega_1 + \omega_2 = 1$. This article quantifies the degree of interest in the user's access to the content. The frequency $F_i^{c_j}$ is defined as (4)

$$F_i^{c_j} = \varepsilon_1 \times \frac{1}{S_i^{c_j}} + \varepsilon_2 \times R_i^{c_j} + \varepsilon_3 \times V_i^{c_j} \tag{4}$$

Where ε_1, ε_2, and ε_3 represent weights, which range from 0 to 1 and satisfy $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = 1$. $S_i^{c_j}$ represents the reading time interval of the user's access content c_j, $R_i^{c_j}$ corresponds to the number of times of reading of the user's access content c_j, and $V_i^{c_j}$ corresponds to the average number of times of reading of the user's access content c_j. This article considers that the shorter the time and the more read times, The user is more interested in the content [11].

Similarly, the viscosity $T_i^{c_j}$ is defined as (5)

$$T_i^{c_j} = \eta_1 \times E_i^{c_j} + \eta_2 \times G_i^{c_j} \tag{5}$$

Where η_1 and η_2 represent weights, which range from 0 to 1 and satisfy $\eta_1 + \eta_2 = 1$. $E_i^{c_j}$ is the user's reading time for the preference c_j, and $G_i^{c_j}$ is the average daily reading length for the user's preference for this class c_j.

Definition 4 A matrix that reflects user preferences[12]. $A = (A_{ij})_{m*n}$ is called the user preference relation matrix, $i = 1, 2, ..., m$ represent the user, and $j = 1, 2, ..., n$ represent the category of user preference. The determination threshold for setting the content preference is $P_{th}^{c_j}$, where A_{ij} is as follows

$$A_{ij} = \begin{cases} T, & P_i^{c_j} \geq P_{th}^{c_j}, \text{ like} \\ F, & P_i^{c_j} < P_{th}^{c_j}, \text{ dislike} \end{cases}$$

Definition 5 Multi - context preference model. The user preference model contains user preferences and contextual information, as in (6)

$$U_{interest(i)} = \{< i, situation(i), interest(i) >\} \tag{6}$$

Where *sitution*(i) is the context information of the user, and *interest*(i) is the preference of the user after the introduction of the context factor, which is obtained by the preference matrix, context factor and Apriori algorithm obtained by definition 4.

2.2 Multi - context user preference analysis

When the mobile user accesses a site, this article classifies the content that the user browses, and obtains the user's interest in accessing the content according to the definition above.

(1)Data preprocessing

Data preprocessing is divided into data collection and extraction preprocessing. But they are not described here, the user access to the content to be divided to improve the efficiency of data analysis.

(2) Analysis Strategy of User Preference in Multi - context

According to the data obtained from the above classification, the user preference matrix is constructed by giving the user access information to give the decision threshold of the content preference. In this paper, the Apriori algorithm in SPSS Modeler is used to mine the user preferences. Specific steps are as follows

Step 1 This article first obtains the mobile user's network log, uses the URL to extract the text information to carry on the classification to visit the content;

Step 2 obtain the current user access to the situation information (through the operator base station data calculation and analysis);

Step 3: Calculate the user's interest in the access preference by definition 3, set the threshold, and get the user preference matrix, which indicates whether the user likes or not. And the matrix is introduced into SPSS Modeler, using the association rule Apriori algorithm to draw the user preference associated mesh map;

Step 4 Set the standard for weak connection in SPSS Modeler, which is a strong link with more than 35% of the total link, and 15% or less for the weak connection;

Step 5 According to the result of Step 4, the users who are not interested in the above preferences and the users with low browsing frequency are selected to match the context information extracted in Step 2.

Step 6 analyzes the user's preferences in different contexts based on the preference model.

Multi-context user preference analysis process shown in Figure 1.

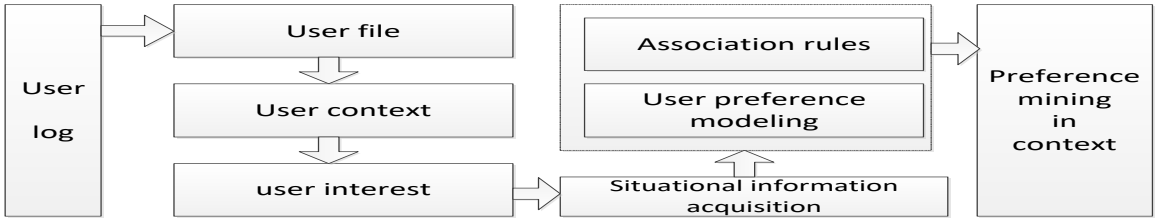

Figure 1 User preference analysis process

Therefore , we can get some preferences which the user is interested in in different situations, record the access time of the mobile user's Internet and the access frequency when the mobile user visits the Web site, get the relationship matrix between the user and its preference. In the steps 3 and 4, you need to use the Apriori algorithm in SPSS Modeler to analyze the user's mainstream preferences. Finally, this article obtains the user situation and combines the interest to get the user's preference content.

3. Examples and verification

In this paper, we select the desolate network log data of mobile users in March and April of a city-level operator, and obtain the browsing data (600,000) of 100 mobile users by sampling.

3.1 Case Analysis

This paper focuses on the analysis of mobile users' browsing information in March, and makes use of April data for comparative analysis. For easy calculation, Geographic information is identified as l_1, l_2, l_3. Time information is divided into the morning, afternoon, evening, early morning.According to the definition 3, and the steps 1, 2 and 4, the association relation of the user preference is shown in Figure 2.

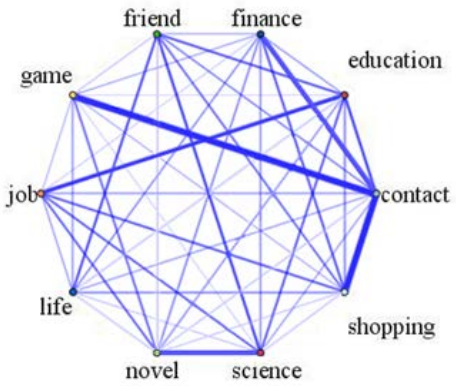

Figure 2 User preference Associated mesh

In Figure 2, the stronger the association between the two preferences, the thicker the line, the weaker the line, the finer the line. We divide the association between the preferences of the user into strong associations, medium associations, and weak associations. Determine the number of connections for the total number of connections more than 35% for strong links, accounting for 15% of the following weak link. From the figure you can see that there are two strong links, namely, contact and shopping, contact and game. Among them, there are three middle links, including novel and science, education and job, finance and contact. This shows that users are most concerned about the process of browsing contact, games and shopping and so on.

According to the definition 3, the preference degree of the user 13 ***** 2341 is shown in Table 2

Table 2 Example of User Context Information

User access to Web site	Preference	Preference interest
c_1	contact	66%
c_2	education	30%
c_3	science	54%
c_4	novel	54%
c_5	life	60%
c_6	game	71%

The table shows that in the general case the user has a high interest in the game preferences, followed by contact and life preferences, The interest in education is low.

3.2 Analysis of results

Through step 5, we further analyze the preference classes of 20 users in different time periods and different regions. Here we get the results of one of the users 13 ***** 2341 as shown in Table 3.

Table 3 Example of User Context Information

Time slot	position information	Preference	frequency	percent
morning	l_3	science	212	49.88%
		education	213	50.12%
afternoon	l_1	science	134	100.00%
	l_2	game	432	100.00%
	l_3	contact	543	81.65%
		science	122	18.34%
evening	l_1	contact	512	100.00%
	l_2	novel	132	100.00%
	l_3	shopping	115	52.27%
		education	105	47.72%
early morning	l_1	science	245	100.00%
	l_3	life	221	61.04%
		novel	141	38.95%

Based on the above analysis results, according to the user in a certain period of time and a region of the browsing record frequency and the proportion of percent,This article believes that the higher the frequency and proportion of browsing, the more inclined the user to the situation.

To verify that the results are correct, we compare the data with the April data and evaluate the results using the similarity defined earlier. According to definition 2, the similarity of 20 users is shown in Figure 3.

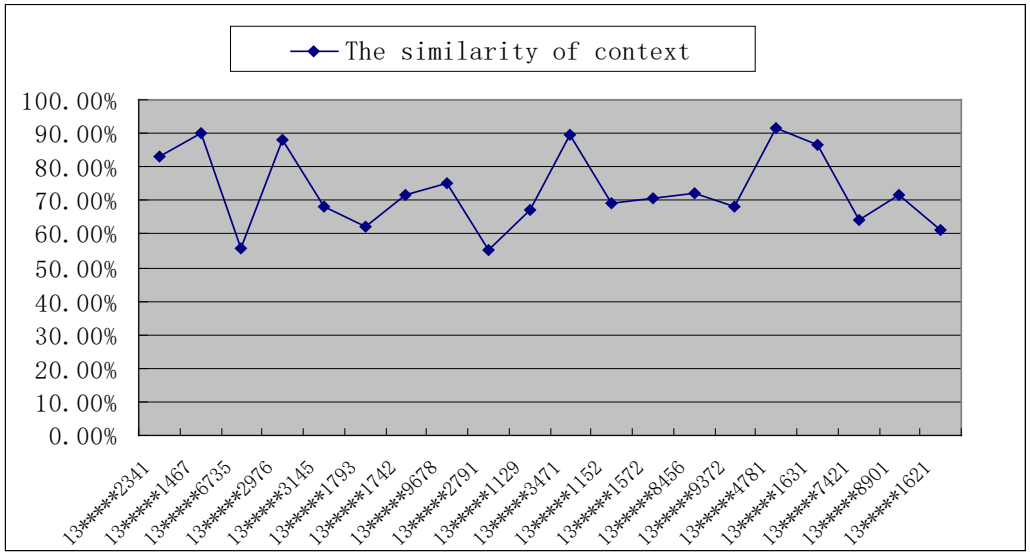

Figure 3 User context similarity

It can be seen from the results in Figure 3, the user context similarity is higher, more than 55%, the average value of 72.7%, indicating that the user will choose a relatively stable time and context of the situation, which to a certain extent that the user For the situation is also preferred.

4. Conclusions

The user's different preferences will be expressed on the Internet for different types of information on the preferences of access differences, we need to face and deal with more complex situations in more cases. This paper presents a multi-context user preference analysis method by calculating the user to browse the time interval and frequency to calculate the mobile users of different preferences of the size of the interest, On this basis,we consider the characteristics of mobile users and combine context information to model mobile users' interest in different contexts.With the deepening of the understanding of smart recommendations under the mobile Internet,the next step in this article will design and develop a more effective model to capture user interest to build and implement personalized recommendation system.

5. Acknowledgments

This work was financially supported by Chongqing University of Posts and Telecommunications Doctoral Foundation (NO.A2015-18).

References

[1] Tveit A. Peer-to-peer based recommendations for mobile commerce[C]//Proceedings of the 1st international workshop on Mobile commerce. ACM, 2001: 26-29.

[2] HU Mu-hai, CAI Shu-qin, ZHANG Yu. Research on Personalized Recommendation-oriented Contextual User Profile[J]. Information Journal, 2010, 29 (10): 157-162.

[3] MENG Xiang-wu, WANG Fan, SHI Yan-cui, et al. Mobile User Requirements Acquisition Techniques and Their Applications[J] .Journal of Software, 2014, 25 (3): 439-456.

[4] Yao X, Shu H. Study on value-added service in mobile telecom based on association rules[C]//Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009. SNPD'09. 10th ACIS International Conference on. IEEE, 2009: 116-119.

[5] Zeng J, Li F, Liu H, et al. A Restaurant Recommender System Based on User Preference and Location in Mobile Environment[C]//Advanced Applied Informatics (IIAI-AAI), 2016 5th IIAI International Congress on. IEEE, 2016: 55-60.

[6] LUO Jun-zhou, WU Wen-jia, YANG Ming.Mobile Internet: Terminal Devices, Networks and Services[J] .Journal of Computers, 2011, 34 (11): 2029-2051.

[7] Whitmore A, Agarwal A, Da Xu L. The Internet of Things—A survey of topics and trends[J]. Information Systems Frontiers, 2015, 17(2): 261-274.

[8] Huang Chuanhui, Sun Yusheng, Ming Chao, et al. Study on the behavior of user's academic information under the guidance of situational user preference[J]. library work and research, 2015, 1 (): 36-39.

[9] Liu Binqiang. Research on personalized recommendation of user interest modeling in mobile environment [D]. Beijing: Beijing University of Posts and Telecommunications, 2009

[10] Huang Guoqing, Wang Mingxu, Wang Guoliang. Weight assignment research of improved entropy method in effectiveness evaluation[J]. computer engineering and application, 2012, 48 (28): 245-248.

[11] Li Jianting, Guo Ye, Tang Zhijun. User interest degree calculation based on analysis users' browsing behaviors [J]. computer engineering and design, 2012, 33 (): 968-972.

[12] Gong Jing, Zhou Feifei, Lu Jia, et al. Analysis of Clients Internet Access Behavior of China Mobile Based on Content Preference[J]. computer technology and development, 2013, 22 (12): 149-152.