PERSONALIZED RECOMMENDATION OF MOBILE TOURISM: A MULTIDIMENSIONAL USER MODEL

Meihua Wei, School of Business, East China University of Science and Technology, Shanghai, China, mei_hua_wei@163.com
Ling Ma, School of Business, East China University of Science and Technology, Shanghai, China, maling@ecust.edu.cn (corresponding author)
Wei Chang, School of Business, East China University of Science and Technology, Shanghai, China, changwei06@126.com

Abstract

With rapid advances in e-business and mobile technology, the personalized recommendation of mobile tourism becomes a critical issue for both researchers and practitioners. The big data, problems of new users and similar recommendations remain barriers for mobile tourism. Through a large dataset gathered by questionnaires, this paper develops a novel multidimensional user model from the perspective of context. The dimensions of our model include several factors: historical behaviour, context and demographic feature of users. To make a better understanding of the model, a case study was adopted. Besides, an experiment is also conducted to evaluate the performance of the proposed model. As a conclusion, limitations and future researches are discussed.

Keywords: mobile tourism, personalized recommendation, user model, context.
1 INTRODUCTION

Technological developments allow us to use IT while we are “on the move”. Since mobile devices permeate deeply into people's life and the desire of tourism is surging, “mobile tourism” has become popular. For example, the online mobile tourism platform was convenient to adopt services (Deng, X. et al. 2011). However, the overloading information in attractions and the unique requirements of individual bring forward new challenges but important opportunities for the evolution of innovative personalized services. And personalized recommendation system of mobile tourism is an important step toward alleviating tourists’ burden (Zhang, T. C. et al. 2011), which is just like an online travel agency for solving various problems of tourists.

The main goal of personalized recommendation is to present the information that individuals want to see in an appropriate manner and time (Pierrakos et al. 2003). Recommendation systems have been used in a wide variety of application domains (e.g. Films, books, news, web pages etc), especially in tourism recently (Kenteris, M. et al. 2010). In mobile tourism, personalization has mainly been addressed to match user preferences, and context-aware researches have mostly focused on user’s physical location, orientation, speed, etc(Noguera, J. M. et al. 2012). The context-aware literature based on location considered tourists’ preferences, needs, and constraints. location, time, destination and attraction information, as well as restaurants requirements(Yu, C. C. and Chang, H. P., 2009). However, they didn’t take users’ demographic features into account. Although the sparsity and cold start problems have been solved partially by using Collaborative Filtering (CF), Content Based Filtering (CBF), clustering and association rules, these methods didn’t consider different context of users (Mohammadnezhad, M., and Mahdavi, M., 2012). Generating high-quality recommendations for new users is a special challenge because of lack of available information (Huang, Z. et al. 2004).

While these researches have provided significant insights into different aspects related to mobile tourism, the extent research is fragmented and has evolved over the years in a fairly nonintegrated way. This paper takes the study of the user model of personalized recommendation a step further in that direction by proposing an integrative model, the user model of mobile tourism, which solves these problems and improve accuracy of recommendation (Ansari, A. et al. 2000).

The main purpose of our study is to improve the quality of personalized recommendation of tourism service. Tourists are a heterogeneous group of customers with different ages, interests, habits etc, and have different tastes, experiences and motivations of travelling. Furthermore, tourism is an activity in different contexts, implying a need to design for “multi-model” (Eriksson, C. I., and Åkesson, M., 2007). Therefore, we develop a multidimensional user model based on context, demographic feature and historical behaviour of users. Through collecting questionnaires and analyzing data, we made an evaluation by comparing the common evaluation metrics for recommender system, i.e. precision, recall, and F-measure.

As such, service providers should consider three basic aspects of tourists comprehensively, namely, context, demographic feature and historical behaviour of users, to gain a large customer base. Besides, the paper provides a global perspective to understanding the desire of tourists. By showing the importance and complexity between users and mobile tourism, the paper calls for more research on the topic.

2 LITERATURE REVIEW

The concept of mobile tourism has recently emerged wherein tourists are able to get information before, during and after their trip with the help of mobile applications (Kenteris, M. et al. 2009). Mobile tourism has a great advantage on flexibility of schedule (El-Sofany, et al. 2011). Mobile Tourist Guide based on location orientation, has provided to improve service of tourism by semantic web (Fluit, C. et al. 2006; Yueh, Y. T. et al. 2007), 3D-GIS (Noguera, J. M. et al. 2012), Google map (Pan, B. et al. 2007). And some researchers tried to find an appropriate application for tourists. However, through analyzing the data collected in three cases, problems of hardware and wireless
connection, the outdate information still exist (da Silva et al. 2012). As the advantages of personalized recommendation for customized tourism, many studies have explored it.

Approaches of recommender systems include collaborative filtering (Goldberg, D. et al. 1992), content-based (Lops, P. et al. 2011), knowledge-based (Burke, R. 2000), and hybrid-based methods (Woermdl, W. et al. 2007). Many of them mostly concerned customized products (Senecal, S., and Nantel, J. 2004), travel schedule (Liu, Q. et al. 2011) and information retrieval (Herlocker, J. L. et al., 1999). There are two research streams in personalized recommendation of mobile tourism. Some researches extended users’ information and constructed user model by analysing tourists’ behaviour and specific attractions (Weng, S. S., Lin, B., and Chen, W. T. 2009; Hinze, A., and Buchanan, G. 2005). Others focused on the limitations of traditional recommended methods, such as cold start, new users, and data sparsity. Collaborative error-reflected models, the integration of EBM model (Kim, H. N., El-Saddik, A. et al. 2011), Bayesian network and Google Maps were proposed respectively (Hsu, F. M. et al. 2012). Although many researches have provided effective models or recommendation systems in tourism, comprehensive user models considering multidimensional attributions both of user and attractions are rare in previous study. And despite all of these achievements of previous studies, their objectives are different with our paper. For example, users’ behaviour model was constructed by expanding users’ features and relative information (García-Crespo, A. et al. 2011), but it is an expert system based on semantic fit. Based on Bayesian network, AHP and behaviour of similar users, a personalized recommendation system was suggested, but it aimed at helping tourists in an unfamiliar city (Huang, Y., and Bian, L. 2009). The purpose of our model is to recommend attractions to tourists rather than travelling schedule. As a result, we did not take route-scheduling factor into account, such as transport, accommodation.

Context is great opportunities in travel and tourism industries to achieve either personalized and interactive products or service recommendation (Cheverst, K. et al. 2002; Liu, L. et al. 2013). Content of context include three perspectives: physical context (Huang, Y., and Bian, L. 2009), service context (George, J. M., and Bettenhausen, K. 1990) and historical context (Gerin, W. et al. 2000). Physical context refers to any information describing the state of entity, such as time, location, tasks, and people in surroundings (Dey, A. K. et al. 2001). It has been acquired by mobile device and was used to recommend adaptive transportation, hotels and route planning for tourists (Noh, H. Y. et al. 2012). Service context consists of condition, state, required resources of task. Namely, anything relates to finish tasks (Kersten, M., and Murphy, G. C. 2006). As the number of resources available to help knowledge workers is increasing, a context-sensitive collaboration system provided information and human resources based on respective goals was developed(Gong, R. et al. 2009). Historical context mainly concerns about users’ experienced context in the past. The patterns of users’ behaviour and relevant preferences can be inferred by analysing historical behaviour (historical context is identical with historical behaviour in our paper). Historical context has been adopted in shopping and catering to achieve better services (Hong, J. et al. 2009). As such, context has been used widely in mobile business, but few of them synthesize the three perspectives of context in one model. If any, they tended to present a novel mobile recommender system with a mobile 3D GIS architecture, map-based interface, real-time location and contextual geofencing mobile tourism service (Noguera, J. M. et al. 2012; Martin, D. et al. 2011). Thus, the contextual content in our research refers to physical context (weather, time); service context (attraction type, tourist type, and travel motivation) and historical context (visited attractions, new attractions and feedback). And we also take demographic features of tourists into account (Adomavicius, G., and Tuzhilin, A. 2005).

3 MULTIDIMENSIONAL USER MODELING TO MOBILE TOURISM BASED ON CONTEXT

3.1 Inference of multidimensional user modeling

According to literature reviews, we create a Bayesian network topology model of preferred attractions. The multidimensional model built on Bayesian network, is shown as figure 1.
Figure 1. Bayesian network model of travel attractions based on current context (left area), demographic features (area below) and historical behaviour of user (right area)

The posterior probability of preferred attraction is defined as equation (1). \( P(\text{PA}|C) \) is the probability of preference for nature scenery, cultural relics, outdoor activities, relaxation, local attraction.

\[
P(\text{PA}|C) = \frac{P(T|\text{PA})P(T_m|\text{PA})P(A_t|\text{PA})P(\text{PA})}{\sum_{\text{PA}} P(C|\text{PA})P(\text{PA})}
\]  

(1)

Where the values of \( j \) are N, C, O, E, L and \( i \) equals cur, hit or b, and \( C \) means current context, historical behaviour and demographic features of users respectively.

3.2 Representation of multidimensional model based on context

Vector is a common personalized recommendation algorithm for similar user Bieliková, M., et al., 2012). This study describes the inference results of Bayesian network inference as \( 1*5, 1*4 \) and \( 1*7 \) vector respectively, as (2).

\[
\begin{align*}
U_{\text{cur}} &= (P_{\text{At}}, P_{\text{PA}}, P_{\text{Tm}}, P_{\text{cur}}, P_{\text{PA}}) \\
U_{\text{hit}} &= (P_{\text{Hv}}, P_{\text{Nv}}, P_{\text{F}}, P_{\text{PA}}) \\
U_{\text{b}} &= (P_{\text{G}}, P_{\text{Age}}, P_{\text{Occu}}, P_{\text{AI}}, P_{\text{Psnty}}, P_{\text{PA}})
\end{align*}
\]  

(2)

where \( 0 \leq P_i \leq 1 \); and \( i = T, At, Tt, Tm, PA, Hv, Nv, F, G, Age, Occu, AI, Edu, Psnty \) correspondingly.

3.3 The weight calculation of dimensions of user model

The probabilities of preferred attractions based on Bayesian network inference under different dimensions are defined as formula (3):

\[
\begin{align*}
U'_{\text{cur}} &= (P_{u_1}, P_{u_2}, P_{u_3}, P_{u_4}, P_{u_5}) \\
U'_{\text{cur}} &= (P_{u_{19}}) = (P_{u_{11}}, P_{u_{12}}, P_{u_{13}}, P_{u_{14}}, P_{u_{15}}) \\
U'_{\text{hit}} &= (P_{u_{20}}) = (P_{u_{1}}, P_{u_2}, P_{u_3}, P_{u_4}, P_{u_5})
\end{align*}
\]  

(3)

where \( 0 \leq P_u \leq 1, 0 \leq P_{u_{i}} \leq 1, K = \text{"cur" or "hit"}. \)

The weight means the effective degree of current context, historical behaviour and demographic
features on users’ preferences, which can be denoted as the vector calculation of correlation coefficient, defined as (4):

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

(4)

Based on formula (4), the correlation coefficient between $U_{\text{cur}}'$ and $U_{\text{hit}}'$, $U_{b}'$ and $U_{\text{cur}}'$, $U_{b}'$ and $U_{\text{hit}}'$ can be calculated by formula (5). And the weights of them represented by $\omega_{\text{cur}}, \omega_{\text{hit}}, \omega_{b}$ respectively, are calculated by formula (6).

$$\rho_{\text{ch}} = \frac{\sum_{i=1}^{n} (P_{\text{cur}} - P_{\text{hit}}) / n - (\sum_{i=1}^{n} P_{\text{cur}} / n)^{2} \cdot (\sum_{j=1}^{n} P_{\text{hit}} / n)}{\sqrt{\sum_{i=1}^{n} (P_{\text{cur}} - P_{\text{hit}})^{2} / n}}$$

$$\rho_{\text{cb}} = \frac{\sum_{i=1}^{n} (P_{\text{cur}} - P_{b}) / n - (\sum_{i=1}^{n} P_{\text{cur}} / n)^{2} \cdot (\sum_{j=1}^{n} P_{b} / n)}{\sqrt{\sum_{i=1}^{n} (P_{\text{cur}} - P_{b})^{2} / n}}$$

where $0 \leq P_{\text{cur}} \leq 1, 0 \leq P_{\text{hit}} \leq 1, i = 1,2,\ldots,5, j = 1,2,\ldots,5.$

$$\omega_{\text{cur}} = (1 + \rho_{\text{ch}} + \rho_{\text{cb}}) / [3 + 2(\rho_{\text{ch}} + \rho_{\text{cb}} + \rho_{\text{hit}})]$$

$$\omega_{\text{hit}} = (1 + \rho_{\text{ch}} + \rho_{\text{hit}}) / [3 + 2(\rho_{\text{ch}} + \rho_{\text{hit}} + \rho_{\text{cb}})]$$

$$\omega_{b} = (1 + \rho_{\text{cb}} + \rho_{\text{hit}}) / [3 + 2(\rho_{\text{cb}} + \rho_{\text{hit}} + \rho_{\text{ch}})]$$

where $0 < \omega_{\text{cur}}, \omega_{\text{hit}}, \omega_{b} < 1.$

(6)

3.4 Principle of personalized attraction recommendation based on context

The recommendation processes include system input, user preference model, personalized attraction recommendation, and system storage. The process of personalized attraction recommendation in mobile tourism based on context is as figure 2.
User groups are divided into existing users group and nonexistent users group. As for existing users, inferred attractions are recommended to users according to their historical records and current demands. However, when the system does not have the basic characteristics and the historical records of the user, recommendation is delivered by matching new users with existing users based on similar context and demographic features. The similarity calculation between users is necessary for users, which makes the recommendation more efficient. Therefore, similar users in database should be identified firstly. According to previous study, cosine-based similarity is used to calculate user similarity (Linden, G. et al., 2003). According to the formula of user characteristics as (7), the similarity calculation based on user preference model is as (8).

$$\hat{U}_a = (\omega_{u_a} \cdot \hat{U}_{cur}, \omega_{u_a} \cdot \hat{U}_{his}, \omega_{u_a} \cdot \hat{U}_{ba})$$
$$\hat{U} = \{\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n\}$$

$$\hat{u}_k = (\omega_{u_k} \cdot \hat{u}_{cur}, \omega_{u_k} \cdot \hat{u}_{his}, \omega_{u_k} \cdot \hat{u}_{ba})$$

Where $U_a$ is current user, $U$ represents existing user in database.

$$\cos(\hat{U}_a, \hat{u}_k) = \frac{\hat{U}_a \cdot \hat{u}_k}{\|\hat{U}_a\| \cdot \|\hat{u}_k\|} =$$

$$\frac{\omega_{u_a} \cdot \omega_{u_k} (T_{T_a} T_{T_k} A_{A_k} T_{A_a} A_{A_a} T_{A_k} T_{A_a} + PA_{cur} PA_{cur}) +}{\|\hat{U}_a\| \cdot \|\hat{u}_k\|}$$

$$\frac{\omega_{u_a} \cdot \omega_{u_k} (H_{H_a} H_{H_k} + N_{N_a} + F_{F_k} + PA_{his} PA_{his}) +}{\|\hat{U}_a\| \cdot \|\hat{u}_k\|}$$

$$\frac{\omega_{u_k} (G_{G_k} + A_{A_k} + O_{cur} O_{cur} + AI_{AI} + E_{desk} E_{desk} + P_{nature} P_{nature} + PA_{ba} PA_{ba})}{\|\hat{U}_a\| \cdot \|\hat{u}_k\|}$$

3.5 A case study

The method is demonstrated by a case study with an existing traveller $U$ in database. The current context can be described: a 23-year-old college girl is going to travel to one nature scenery to relax. As a college student in school, her living expense is 1500 yuan a month. It is a sunny day with comfortable breeze in the morning. According to the system database (all initial data is obtained by questionnaire and calculated by probability formula in this paper), she once enjoyed some relaxation attractions, but she has never selected cultural relics. Using the professional software of Bayesian network inference -Netica, the results based on current context are showed as figure 3. (Netica principle of demographic features and historical behaviour is identical, except inferred values)

![Figure 3. The result of Bayesian Network inference based on current context](image)

According to the section 3, results of preferred attractions were explained by equation (10).

$$\hat{U}_{cur} = (0.1806, 0.0920, 0.0770, 0.3456, 0.3048); \quad \hat{U}_{his} = (0.0438, 0.0799, 0.4976, 0.2027, 0.1760)$$

$$\hat{U}_b = (0.0434, 0.0298, 0.4574, 0.0809, 0.3885)$$

(10)
The weights are $\omega_{\text{usu}} = 0.1694$, $\omega_{\text{u}} = 0.4387$, $\omega_{\text{usu}} = 0.3919$ respectively. Thus, multidimensional user model of mobile tourism can be explained by formula (11):

$$
\tilde{U} = (0.1694U_{\text{usu}}, 0.3919U_{\text{u}}, 0.4387U_{\text{b}})
$$

$$
\tilde{U}_{\text{usu}} = (0.5302, 0.5025), \tilde{U_{\text{u}}} = (0.7651, 0.8725), \tilde{U}_{\text{b}} = (0.3456, 0.4243, 0.4574)
$$

(11)

According to the formula (8), the results of similarity can be seen in table 1, and table 2 presents the similar user model of traveller U.

<table>
<thead>
<tr>
<th>Sim(U,U1)</th>
<th>Sim(U,U2)</th>
<th>Sim(U,U3)</th>
<th>Sim(U,U4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8639</td>
<td>0.8563</td>
<td>0.8499</td>
<td>0.7835</td>
</tr>
</tbody>
</table>

Table 1. Similar users in top four

From the table 1, the similarity between traveller U1 and U is 0.8639 with the largest value among users in record. U2 is secondary, followed by U3. Therefore, traveller U1, U2 and U3 were selected as similar users for the college student. The maximum probability of attributions in each dimension provides enough evidence to select recommended attraction type to users. According to table 2 and travelling record of similar users in database, Shanghai Wildlife Park (recommended by U1 and U3), Zhujiajiao Ancient Waterside (recommended by U2) and Happy Valley in Shanghai (recommended by U4) are recommended to traveller U.

4 EXPERIMENT

To demonstrate the effectiveness of our model, we first construct the datasets through three methods respectively, and apply recall, precision and F-measure to these datasets for comparative purposes.

4.1 Experimental settings

Based on literature review, we choose a mobile context-aware information system that proposes attractions for tourists depending on their location, users’ travelling history and two personal profiles describing interest based on semantic group. The heart of the system is a filter engine coupled with the location engine (Hinze, A., and Buchanan, G. 2005). To make a better comparison with our model, we selected this mobile personalized recommendation system provided by Hinze as a main context-aware approach for its similarities to our model. In addition, we also compared the two methods with collaborative filtering (CF) because it is a traditional but effective method in recommendation systems.
The common evaluation metrics for recommender systems, i.e., recall, precision and F-measure criteria are used to compare the results of recommender systems based on three methods (Sarwar, B. et al., 2000). Datasets are often classified into two sets. Training sets are for recommendation lists with two third of population, while test set with the rest of population are used to calculate precision and recall (Basu, C. et al., 1998). In most cases, high recall will cause low precision. Therefore, F-measure is to evaluate the recommendation system (Pazzani, M., and Billsus, D. 1997). It is a combination of precision and recall criteria as shown in Formula 14:

\[
\text{Precision} = \frac{|\text{Test set} \cap \text{Recommended tours set}|}{|\text{Test set}|}, \quad \text{Recall} = \frac{|\text{Test set} \cap \text{Recommended tours set}|}{|\text{Recommended tours set}|}
\]

\[
F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

(14)

4.2 Experimental procedure

The proposed model of our paper, a mobile context-aware information system provided by Hinze and CF recommend attractions to tourists in the three training sets. Thus, there will be three training sets and three test sets based on three different methods respectively.

The sample of training sets is from professional research website (taidu8.com). Participants will get virtual monetary or bonus points. We intend to collect a total of more than 270 valid responses. About a total of two-third questionnaires are for the three training sets averagely. And the rest of responses are used for three test sets averagely. All users in our selected sample are able to use mobile device to surf internet and have travelling experiences.

We set one context in the front of all questionnaires to make recommended attractions both accurate and specific. All the questionnaires are distributed online again but attraction lists are suggested to these participants based on three different methods accordingly. Also, the responses of same users in training sets are invalid and removed from our data. Through analysing the collected data, we will compare the F-measure value to show that the performance of our model. The larger the value of F-measure represents the better performance of methods.

5 CONCLUSION

By incorporating the different content and weight of contextual dimensions into the recommendation process, we propose a personalized recommendation method of attractions in mobile tourism and conduct an experiment to test our method. Context information, historical context and demographic features were taken into account to build a multidimensional user model. Meanwhile, attractions can be recommended to new users by matching current context and historical context. To make our method clearer, we do a case study to describe the process of our recommendation. Through experiment, the recommendation quality of our model, traditional CF and a context-aware approach are compared. Our model offers an integrating perspective to the study of mobile tourism, which gives service providers a comprehensive understanding of users.

However, this study also has some limitations that remain to be improved in the future research. Firstly, conditional probability is only collected through questionnaire. Future study can adopt other methods, like expert interview. Secondly, only taking attraction as an example imposes a limit on its general application, other services applying our model can be taken into account. Thirdly, and technical implementation should be studied in the future.

Acknowledgments

Funding for this project was provided by the National Natural Science Foundation of China (71001037).
References


