Research Article

Tourist Behavior Pattern Mining Model Based on Context

Dong-sheng Liu and Shu-jiang Fan

*College of Computer Science & Information Engineering, Zhejiang Gongshang University,
Xiasha University Town, Hangzhou 310018, China*

Correspondence should be addressed to Dong-sheng Liu; lds1118@163.com

Received 11 May 2013; Revised 15 September 2013; Accepted 20 September 2013

Academic Editor: Tinggui Chen

Copyright © 2013 D.-s. Liu and S.-j. Fan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Personalized travel experience and service of tourist has been a hot topic research in the tourism service supply chain. In this paper, we take the context into consideration and propose an analyzed method to the tourist based on the context: firstly, we analyze the context which influences the tourist behavior patterns, select the main context factors, and construct the tourist behavior pattern model based on it; then, we calculate the interest degree of the tourist behavior pattern and mine out the rules with high interest degree with the association rule algorithm; we can make some recommendations to the tourist with better personalized travelling experience and services. At last, we make an experiment to show the feasibility and effectiveness of our method.

1. Introduction

With the development of economy and the improvement of people's living standard, more and more people pay more attention to the quality of personalized travelling experience and service. In recent years, there has emerged more and more personalized ways to travel in tourism, such as FIT travel and independent travel. The traditional mode of travel service limits the diversity of service options, and it cannot fully meet the personalized needs of tourists. How to find the laws and the features of the tourist behavior through mining tourist behavior patterns and offer them better services has been a problem in the tourism service supply chain.

There are many researches concentrating on the tourist behavior pattern. Qing analyzed the characteristics of tourism services and the structural properties, constituent elements, and operation mechanism of tourism service supply chain in the context of modern information technology, and he put forward a new tourism service supply chain conceptual model based on tourist personalized demand [1]. Farmaki took the Troodos (Cyprus) as a case to research on the tourist motivation [2]; Martin and Witt proposed tourism demand forecasting model to represent tourists' cost of living

[3]; Smallman and Moore studied on the tourists' decision making [4]; Kim et al. has worked on the Japanese tourists' shopping preference with the decision tree analysis method [5].

These studies only analyzed the tourist from the view point of the psychology and behavioral science of the tourist and have not considered the context set which will influence the tourist behavior patterns. So, in this paper, we take the context into consideration and propose an analyzed method to the tourist based on context to find out the relationship between services in the travel and the context and analyse the important contexts which will influence the tourist behavior. To mine out rules with high interest degree with the association rule algorithm and do some recommendations to the tourist with better personalized travelling experience and services, we propose a method based on network diagram, and it can reflect the relationship of the contexts which influence the tourist behaviour clearly. Through this method, we can delete the low interest degree of tourist behavior patterns; then, we use the Apriori algorithm to mine the association rules of tourist behavior which have high interest degree. Finally, we take an experiment to show the feasibility and effectiveness of our method.

FIGURE 1: Context spectrum.

2. Related Works

2.1. Context. There are many definitions on the context and many researchers work on it. Schilit et al. defined the context as identifications and change of location, people, and objects around them [6]. Brown et al. thought that the context should be defined as the symbols around people or other objects such as location, time, season, temperature, and so on [7]. In paper [8], the definition of context would be extended to the feature information of some objects' situation, such as people, location, and so on. Snowden and Grasso defined the context as the multilevel structure, mainly including the individual layer, the project layer, the group layer, and the organization layer [9]; Gu thought that the context would respond to the transformation based on the computers which are used as the centers to the people; in fact, he defined the context as a spectrum in his paper, as shown in Figure 1. He divided the context into computing context (such as communication bandwidth), user context (such as location), physical context (such as weather, temperature), time context (such as hour), and social context (such as law) [10].

In this paper, we think that the context is the influence factors of the tourist behavior pattern; different contexts will lead the tourist to different behavior patterns. We may take the following contexts into consideration: user, location, time, and device, and service type.

2.2. Association Rule and Apriori Algorithm. There are many association rule algorithms, and these algorithms can be divided into two classes: the first one is mainly focused on improving the analytical efficiency of the association rules; the other one pays more attention to the application of association rule algorithm and how to deal with value type variables and promotes the association of the single concept

layer to multiple concept layers include and further reveals the inner structure of objects.

Apriori algorithm is one of the classical association rule algorithms; the earliest Apriori algorithm was proposed by Agrawal et al. [11]. The algorithm mainly including two parts: producing frequent item sets and producing association rules according to the frequent item sets. The algorithm scans data base, accumulates each item count, collects the items which meet the minimum support ($\min_sup$), finds out the frequent 1-itemsets, and named it L_1 . Then, the algorithm uses L_1 to find out the frequent 2-item sets L_2 and uses L_2 to find out the frequent 2-item sets L_3 and so on and keeps doing these until it cannot find out the frequent k -item sets. In these frequent item sets, it will be defined as a strong-association rule if it reaches the minimum confidence [12]. Since the association rule algorithm was proposed, it has been improved and applied in many fields. For example, Kang et al. applied the association rule algorithm in the Smart home [13], and Zhang et al. used the improved association rule algorithm in the university teaching managements [14].

3. Modeling and Mining Method for Tourist Behavior Pattern Based on Context

3.1. The Context Influence Factors Analysis of Tourist Behavior Pattern. We can consider a tourist as a mobile customer because the tourist moved anytime and anywhere. Presently, there are only a few researchers who work on the mobile customer behavior pattern. Tseng and Lin thought that the service and location are the influence factors of customer behavior in mobile service environment; they proposed a method named SMAP-Mine to mine customer behaviors [15]. Ma et al. took the time context into consideration and constructed a temporal sequence mobile access patterns mining model based on context awareness [16]. Chen et al. studied in the terms of the problem of mining matching mobile access patterns based on joining the following four kinds of characteristics: user, location, time, and service [17]. So in this paper, we think that the context influence factors of mobile customer behavior pattern includes mobile user, location, time, and service type.

At the same time, we take different capabilities of the mobile devices that the customer use, such as screen size, battery durability, and access bandwidth, into consideration. We consider that these capabilities will influence the mobile customer behavior pattern directly or indirectly. To prove that, we make an experiment as follows. In the particular context, we observed behavior patterns of three customers who used different equipments and recorded the service types, the trajectory at which they moved, and time and type of service. Finally we got the customer movement trajectories as shown in Figure 2 and the service request information table as shown in Table 1. We can conclude from Figure 2 that customers have different behavior patterns when they use different mobile devices. For example, when the user u_1 used the device d_1 , his movement trajectory was $l_2 \rightarrow l_6 \rightarrow l_8 \rightarrow l_9$; when he used the device d_2 , his movement trajectory changed to $l_2 \rightarrow l_6 \rightarrow l_9$. Then, we can conclude from

TABLE 1: Customer service information table.

Time instances	Users and devices					
	(u_1, d_1)	(u_2, d_1)	(u_3, d_1)	(u_1, d_2)	(u_2, d_2)	(u_3, d_2)
TI ₁	(l_2, t_1, s_1)	(l_9, t_5, s_1)	(l_5, t_{13}, s_1)	(l_2, t_1, s_1)	(l_9, t_1, s_1)	(l_5, t_1, s_1)
TI ₂	(l_6, t_2, s_2)	(l_4, t_6, s_2)	(l_6, t_{14}, s_3)	(l_6, t_2, s_3)	(l_4, t_4, s_2)	(l_6, t_4, s_3)
TI ₃	(l_8, t_3, s_3)	(l_1, t_8, s_2)	(l_4, t_{15}, s_2)	(l_9, t_3, s_4)	(l_7, t_5, s_4)	(l_3, t_5, s_2)
TI ₄	(l_9, t_4, s_4)	(l_9, t_9, s_3)	(l_3, t_{16}, s_4)		(l_1, t_6, s_5)	(l_9, t_7, s_6)
TI ₅		(l_7, t_{10}, s_4)	(l_9, t_{17}, s_6)		(l_3, t_8, s_6)	
TI ₆		(l_1, t_{11}, s_5)				
TI ₇		(l_3, t_{12}, s_6)

FIGURE 2: Movement trajectories of customers when they use different devices.

Table 1 that the customer requested different services when he used different devices in the same time or requested the same service in different times. For example, when the user u_1 used device d_1 at time t_3 ; he requested the service s_3 ; when he used device d_2 at time t_3 , he requested the service s_4 ; the user u_3 requested the service s_2 at the location l_4 when he used device d_1 ; he requested the service s_2 at the location l_3 when he used device d_2 . Through these analyses, we can conclude that the mobile customer has different movement trajectories, request different services at the same times and requests the same service in different places when he or she uses different devices. So we take the mobile device as a context influence factor of mobile customer behavior pattern.

There are other context factors which influence the mobile customer behavior pattern, such as the physically environmental condition in which the customer stays, including weather, temperature, humidity, and so on; and the social situations in which the customer is involved (e.g., manners and customs and laws) will influence the mobile customer behavior pattern.

We use the form of the questionnaire to determine the main context factors. In this questionnaire, we design nine questions. Each of the nine questions involves a context factor which will influence the tourist behavior pattern. From these questions, we can study which contexts will influence

FIGURE 3: The results of the questionnaire.

the tourist behavior pattern most. A total of 102 individuals participate in the survey; they are all tourists. After stating these questionnaires, we use SPSS to analyze the results. We set that different option to different weight (1–5), and then statistically averaging, what are the context weights influence the behavior. We can get the results as shown in Figure 3. So in this paper, we choose the following five context factors as the main context factors: tourist (user), device, location, time, and service.

3.2. Modeling the Tourist Behavior Pattern Based on Context.

The preceding part of this paper has a brief analysis on the context factors which influence the tourist behavior pattern, and then we will build a model based on these context factors. In the following part of this paper, we will give relational definitions about the tourist behavior patterns firstly and construct a model of the tourist behavior pattern based on context latterly.

Definition 1 (tourist user). $U = \{u_1, u_2, u_3, \dots, u_i\}$ is the set of all the users; every user denotes a person who uses the mobile device to request mobile service messages from the mobile service supplier when he or she was travelling.

Definition 2 (devices of the tourist use). The device of the user use is a set of the devices of the user use to request mobile services and can be defined as $D = \{d_1, d_2, d_3, \dots, d_h\}$.

TABLE 2: Timestamp table.

Timestamps	Time intervals
t_1	0:00–1:00
t_2	1:00–2:00
t_3	2:00–3:00
$\vdots$	$\vdots$
t_{22}	21:00–22:00
t_{23}	22:00–23:00
t_{24}	23:00–24:00

Definition 3 (location). Location denotes a set of places in which the tourist moves some times, and we use the set $L = \{l_1, l_2, l_3, \dots, l_j\}$ to represent it.

Definition 4 (service). Service is a set of messages in which the tourist requests tourism services from the suppliers, and we use the set $S = \{s_1, s_2, s_3, \dots, s_n\}$ to represent it.

Definition 5 (timestamp, sojourn time and service request time). To represent the time quantum of the forming of the tourist behavior pattern approximately, this paper divides a day's 24 hours into 24 time intervals simply, as shown in Table 2; every time interval denotes one hour, and the hour denotes one timestamp; sojourn time t_s denotes the time in which the user sojourns at somewhere; service request time t_r denotes the time in which the tourist requests some tourism services.

According to the previous definitions, this paper assumes $p = \{u, d, t, l, t_s, s, t_r\}$ as one tourist behavior, where u is an element of the tourist user set U , d is an element of the device of the user use set D , t is an element of the time set T , l is an element of the location set L , t_s is the time in which the tourist sojourns at location l , s denotes an element of service messages set S , and t_r denotes the time in which the tourist requests for tourism services.

In the graph theory, there is a structure called network whose structure is composed of nodes and edges. Every edge has its quantitative index related to the nodes or edges; this quantitative index is normally called weight which could denote distance, expenses, carrying capacity, and so on [18]. Namely, the structure of the network is composed of nodes and edges involving weight; taking this advantage of the network, this paper makes the context factors which influence the tourist behavior pattern as the nodes of the network, the connected relationship among the context factors as the edge of the network, and the connect coefficient among different context factors as the weight of the edge (the specific connect relationship and the connect coefficient will be demonstrated in detail in the following part of this paper). Like this, the behavior pattern of a tourist can be clearly portrayed in the network. Figure 4 illustrates the network structure of the behavior patterns of two different mobile users; we use $p_1 = \{u_1, d_1, l_1, 4, s_1, 2\}$ and $p_2 = \{u_2, d_2, l_2, 1, s_1, 3\}$ to represent their behavior patterns, respectively.

FIGURE 4: The network model of the tourist behavior pattern.

3.3. Tourist Behavior Pattern Mining Method Based on the Network. The preceding part of this paper has a model analysis on the structure of the network of the tourist behavior; in the following part of this paper, we will give out the related definitions firstly and the specific procedures of the tourist behavior mining pattern based on the network latterly.

3.3.1. Basic Definitions. To explain the content of the mining method more clearly, we will give relational definitions firstly.

Definition 6 (connect coefficient). Connect coefficient denotes the connection relationship between two different attributes; the specific connect coefficients are $U \bowtie D$, $D \bowtie L$, $L \bowtie T$, and $T \bowtie S$. The connect coefficient of $U \bowtie D$ denotes the connection times between a mobile user u and a device d . The connect coefficient of $D \bowtie L$ is LC_{nu} which denotes the connection times of a device d with a location l . The connect coefficient of $L \bowtie T$ is t_s which denotes the time in which a user sojourns at location l . The connect coefficient of $T \bowtie S$ is t_r which denotes the time in which a mobile user requests for services.

Definition 7 (interesting locations and interesting services). When the length of time in which a tourist sojourns somewhere is larger than the threshold value we set, we think that the tourist is interested in this place. Similarly, when the length of time in which a mobile user requests for a service s_j is larger than the threshold value or the connection times is larger than a threshold value, we think that the mobile user is interested in this service. Usually the length of time will be set up to 30 minutes and the connection times will be set up to 10 times.

Definition 8 (repeated edge). For a tourist, he may have the same connection edge in two different behavior patterns; such edge will be called repeated edges in this paper. For example, in the following behavior patterns $p_1 = \{u_1, d_1, l_3, t_{17}, s_3\}$ and $p_2 = \{u_1, d_2, l_3, t_{17}, s_3\}$, they have two repeated edges, namely, $l_3 t_{17}$ and $t_{17} s_3$.

Definition 9 (connect edge value). Connect edge value is a standard value obtained with standardizing the connect coefficient (Definition 6) in the case where the different quantity levels of input variables affect the final mining result. In this paper we use " $\bowtie$ " to present the connect relationship between different attributes, and specific weights are $U \bowtie D$,

$D \bowtie L$, $L \bowtie T$, and $T \bowtie S$; the computational formulas of every edge weight are as follows.

Connect Edge Value of $U \bowtie D$. The connect edge value of mobile user u_i and device d_j equals the ratio of the connect times between user u_i and device d_j to the sum times between user u_i and device set; the specific formula is

$$w_{ij} = \frac{\text{count}(u_i d_j)}{\sum_{j=1}^n \text{count}(u_i d_j)}, \quad (1)$$

where n denotes the amount of devices. Similarly, the connect edge value of $D \bowtie L$ is as follows:

$$w_{jk} = \frac{LC_{nu}(jk)}{\sum_{k=1}^m LC_{nu}(jk)}, \quad (2)$$

where m denotes the amount of locations and $LC_{nu}(jk)$ and $\sum_{k=1}^m LC_{nu}(jk)$ denote the connect times between devices and locations in the same behavior pattern of a mobile user.

The Connect Edge Value of $L \bowtie T$

$$w_{kh} = \frac{t_s(kh)}{\sum_{h=1}^{24} t_s(kh)}, \quad (3)$$

where $t_s(kh)$ and $\sum_{h=1}^{24} t_s(kh)$ denote the time in which a mobile user requests services at somewhere in his behavior pattern.

The Connect Edge Value of $T \bowtie S$

$$w_{hz} = \frac{t_r(hz)}{\sum_{z=1}^n t_r(hz)}, \quad (4)$$

where n denotes the amount of the connect service set and $t_r(hz)$ and $\sum_{z=1}^n t_r(hz)$ denote the time in which a mobile user requests for services in his behavior pattern. An edge will be deleted if its connection edge value is smaller than a threshold value. A behavior pattern will not be involved in the calculation of the connect edge value if it contains interesting locations or interesting services.

Definition 10 (connect edge coefficient e). When a repeated edge appears, this edge value constitutes of several behavior patterns; connection edge coefficient e denotes the incidence a behavior pattern has on this edge. Its value equals ratio of the connect edge coefficient of this behavior pattern to the sum of all the connect edge coefficients of the same mobile user at this edge.

Definition 11 (interesting degree id). Interesting degree id is an index to reflect the degrees of interests of the mobile user behavior pattern. Specifically, it equals the value that the sum of all the tuple (u, d, t, l, s) weight, the formula of interesting degree id = $e_{ij} * w_{ij} + e_{jk} * w_{jk} + e_{kh} * w_{kh} + e_{hz} * w_{hz}$. If the value of interesting degree id is smaller than a threshold value th_1 , we will regard the degree of interests of this mobile user pattern as low interest level and delete this pattern from

FIGURE 5: Two thresholds and three degrees.

TABLE 3: Tourist information.

U_i	d_h	L_j	LC_{nu}	T_m	T_s	S_n	T_r
u_1	d_1	l_1	1	t_{15}	8	s_4	6
u_1	d_2	l_1	1	t_{15}	8	s_2	2
u_2	d_1	l_4	2	t_{16}	9	s_3	9
u_2	d_2	l_3	2	t_{17}	9	s_3	9
u_3	d_1	l_5	6	t_{17}	5	s_5	5
u_3	d_2	l_5	6	t_{18}	12	s_5	8
u_4	d_3	l_7	5	t_{14}	8	s_6	7
u_5	d_2	l_6	3	t_{16}	10	s_8	9
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

the network. If the value of interesting degree id is larger than another threshold value th_2 , we will regard the degree of interests of this mobile user pattern as high interest level. Like this, we divide mobile user behavior patterns into three parts, namely, low level of interest, common level of interest, and high level of interest. We can set them by our need; the larger the value is, the higher degree of interest the rules of the results will have. As is illustrated in Figure 5, we can use it as a behavior prediction model to predict the behavior pattern of a mobile user in the future. If a behavior pattern contains interesting locations or interesting services, we will regard it as the high interesting level behavior pattern without calculating the specific value of its interestingness.

3.3.2. Mining Steps

First Step (collecting data). To mine tourist behavior pattern, we must collect data about the tourist. We can get the information table as is shown in Table 3 through collecting user data, mainly including tourist information (U_i), mobile device (d_h), location (L_j), collecting times (LC_{nu}), time (T_m), time of the user stay the location (T_s), the service type the user request (S_n), and time of the user request the service (T_r).

Second Step. Extracting context attribute number of the context set which influences the mobile customer behavior pattern and design corresponding layers of the network diagram; in this paper, we should design a network diagram with five layers, each layer corresponds to all nodes of U_i , D_h , L_j , T_m , and S_n , respectively, and the number of the layer nodes corresponds to each attribute value number, as shown in Figure 6.

Third Step. Collecting the adjacent nodes, each connection coefficient should be marked as Definition 7; we need to add the connection coefficient of the side when it repeats several

FIGURE 6: Network diagram model.

times. For example, there are two situations when the device d_1 collects the location l_3 : one is 4 and the other is 2, so the connection coefficient of d_1l_3 equals $4 + 2 = 6$.

Forth Step. Considering different customers have different behavior patterns, we classify each user into a group and calculate the collection weight according to Definition 8; when the collection weight is lesser than the threshold, the edge will be deleted.

Fifth Step. Calculating the remaining customer interest degree according to Definition 11 and set the low interest degree th_1 and the high interest degree th_2 value. When the customer interest degree is lesser than the low interest degree th_1 , this customer behavior pattern will be unconcerned, and the general interest degree and the high interest degree pattern will be conducted in the next step.

Sixth Step. Using Apriori algorithm to mine the frequent pattern to the general interest degree and the high interest degree pattern, mine out the association rules with higher degree value on support and confidence; we can use these rules to forecast the customer's behaviors in future or recommend some services to mobile customers.

In order to show the availability of our method, we propose the concept of "coverage," which means the ratio of the number of the same rules that are produced by our model to the number of rules that produced directly. If the coverage is larger than a threshold, we say that the method we proposed is available. Generally, the larger the threshold is, the more the availability of the method is. In this paper, we set the threshold to be equal to 80%.

4. Experiment and Analysis

4.1. Example and Analysis. We take the West Lake of Hangzhou, for example, to illustrate the application of the model, via GPS and RFID provide personalized services to users combined with requirements and preference of the user. So we select part of the information data about tourist behavior from West Lake of Hangzhou Scenic Area Management Committee as is shown in Table 4.

To verify the effects of the proposed method, we use two standard metrics: interest degree and coverage.

TABLE 4: Tourist behavior information table.

P	U_i	d_h	L_j	LC_{mu}	T_m	T_s	S_n	T_r
p_1	u_1	d_1	l_1	1	t_{15}	8	s_4	6
p_2	u_1	d_2	l_1	1	t_{15}	8	s_2	2
p_3	u_1	d_1	l_1	1	t_{16}	11	s_2	3
p_4	u_1	d_2	l_1	1	t_{16}	11	s_2	4
p_5	u_1	d_1	l_3	4	t_{17}	21	s_3	17
p_6	u_1	d_2	l_3	4	t_{17}	4	s_3	3
p_7	u_1	d_1	l_2	11	t_{16}	35	s_4	30
p_8	u_2	d_1	l_4	2	t_{16}	9	s_3	9
p_9	u_2	d_2	l_3	2	t_{17}	9	s_3	9
p_{10}	u_2	d_1	l_3	2	t_{17}	2	s_3	1
p_{11}	u_2	d_2	l_2	2	t_{17}	28	s_5	15
p_{12}	u_2	d_2	l_2	8	t_{16}	38	s_4	31
p_{13}	u_3	d_1	l_5	6	t_{17}	5	s_5	5
p_{14}	u_3	d_2	l_5	6	t_{18}	12	s_5	8
p_{15}	u_3	d_1	l_5	6	t_{18}	12	s_4	4
p_{16}	u_3	d_1	l_5	6	t_{18}	18	s_3	13
p_{17}	u_3	d_2	l_5	6	t_{18}	18	s_3	5
p_{18}	u_3	d_2	l_5	10	t_{16}	32	s_4	30
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

We can conclude from Table 4 that the patterns p_7 , p_{12} , and p_{18} are the patterns with interested locations or interested services, and we think that these patterns are the behavior patterns with high interest degree patterns. To show the processes of our method, we choose three users' patterns in this paper and design a network diagram with five layers and collect the adjacent layers and then calculate the connection coefficients according to Definition 7, as shown in Figure 7.

Then, we divide each user into a group and calculate the collection weight according to Definition 8; we set that the edges whose weight is lesser than 0.2 would be deleted, so the edges d_1l_1 , d_2l_1 , $t_{15}s_2$, $t_{15}s_2$, $t_{15}s_4$, l_4t_{15} , l_5t_{17} , $t_{17}s_5$, and $t_{18}s_4$ will be deleted and the patterns with edges d_1l_1 , d_2l_1 , $t_{15}s_2$, $t_{15}s_2$, $t_{15}s_4$, l_4t_{15} , l_5t_{17} , $t_{17}s_5$, and $t_{18}s_4$ will be deleted too, as shown in Figures 8, 9, 10.

So remain following patterns: $p_5 = \{u_1, d_1, l_3, t_{17}, s_3\}$, $p_6 = \{u_1, d_2, l_3, t_{17}, s_3\}$, $p_{10} = \{u_2, d_1, l_3, t_{17}, s_3\}$, $p_{11} = \{u_2, d_2, l_2, t_{17}, s_5\}$, $p_{14} = \{u_3, d_2, l_5, t_{18}, s_5\}$, $p_{16} = \{u_3, d_1, l_5, t_{18}, s_3\}$, $p_{17} = \{u_3, d_1, l_5, t_{18}, s_3\}$; then we calculate the interesting degree of these patterns according to Definition 11, as is shown in the following expressions:

$$\begin{aligned}
 id_{p_5} &= e_{11} * w_{u_1d_1} + e_{13} * w_{d_1l_3} \\
 &\quad + e_{3-17} * w_{l_3t_{17}} + e_{17-3} * w_{t_{17}s_3} \\
 &= \frac{1}{3} * 0.5 + 1 * 0.333 + \frac{21}{25} * 0.397 \\
 &\quad + \frac{17}{20} * 0.571 \approx 1.305;
 \end{aligned}$$

$$\begin{aligned}
 id_{p_6} &= e_{12} * w_{u_1d_2} + e_{23} * w_{d_2l_3} \\
 &\quad + e_{3-17} * w_{l_3t_{17}} + e_{17-3} * w_{t_{17}s_3}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{3} * 0.5 + 1 * 0.333 + \frac{4}{25} * 0.397 \\
&\quad + \frac{3}{20} * 0.571 \approx 0.647; \\
\text{id}_{p_9} &= e_{22} * w_{u_2 d_2} + e_{23} * w_{d_2 l_3} \\
&\quad + e_{3-17} * w_{l_3 t_{17}} + e_{17-3} * w_{t_{17} s_3} \\
&= 0.5 * \frac{1}{2} + 1 * 0.25 + 0.229 * \frac{9}{11} \\
&\quad + 0.294 * \frac{9}{10} \approx 0.952; \\
\text{id}_{p_{10}} &= e_{21} * w_{u_2 d_1} + e_{13} * w_{d_1 l_3} \\
&\quad + e_{3-17} * w_{l_3 t_{17}} + e_{17-3} * w_{t_{17} s_3} \\
&= 0.5 * \frac{1}{2} + 1 * 0.25 + 0.229 * \frac{2}{11} \\
&\quad + 0.294 * \frac{1}{10} \approx 0.571; \\
\text{id}_{p_{11}} &= e_{22} * w_{u_2 d_2} + e_{22} * w_{d_2 l_2} \\
&\quad + e_{2-17} * w_{l_2 t_{17}} + e_{17-5} * w_{t_{17} s_5} \\
&= 0.5 * \frac{1}{2} + 1 * 0.25 + 1 * 0.583 \\
&\quad + 1 * 0.441 \approx 1.524; \\
\text{id}_{p_{14}} &= e_{3-2} * w_{u_3 d_2} + e_{25} * w_{d_2 l_5} \\
&\quad + e_{5-18} * w_{l_5 t_{18}} + e_{18-5} * w_{t_{18} s_5} \\
&= 0.4 * \frac{1}{2} + 0.4 * \frac{1}{2} + 0.923 * \frac{12}{60} \\
&\quad + 1 * 0.229 \approx 0.815; \\
\text{id}_{p_{16}} &= e_{31} * w_{u_3 d_1} + e_{15} * w_{d_1 l_5} \\
&\quad + e_{5-18} * w_{l_5 t_{18}} + e_{18-3} * w_{t_{18} s_3} \\
&= 0.6 * \frac{1}{3} + 0.6 * \frac{1}{3} + 0.923 * \frac{18}{60} \\
&\quad + 0.514 * \frac{13}{18} \approx 1.05; \\
\text{id}_{p_{17}} &= e_{31} * w_{u_3 d_1} + e_{15} * w_{d_1 l_5} \\
&\quad + e_{5-18} * w_{l_5 t_{18}} + e_{18-3} * w_{t_{18} s_3} \\
&= 0.4 * \frac{1}{2} + 0.4 * \frac{1}{2} + 0.923 * \frac{18}{60} \\
&\quad + 0.514 * \frac{5}{18} \approx 0.820.
\end{aligned}
\tag{5}$$

In this paper, we set the low interesting degree th_1 value to be equal to 0.8 and the high interesting degree th_2 value

TABLE 5: Patterns with high interesting degree.

P_i	U_i	d_h	L_j	T_m	S_n
p_5	u_1	d_1	l_3	t_{17}	s_3
p_7	u_1	d_1	l_2	t_{16}	s_4
p_9	u_2	d_2	l_3	t_{17}	s_3
p_{11}	u_2	d_2	l_2	t_{17}	s_5
p_{12}	u_2	d_2	l_2	t_{16}	s_4
p_{14}	u_3	d_2	l_5	t_{18}	s_5
p_{16}	u_3	d_1	l_5	t_{18}	s_3
p_{17}	u_3	d_2	l_5	t_{18}	s_3
p_{18}	u_3	d_2	l_5	t_{16}	s_4

to be equal to 1. So the patterns whose interesting degrees are lesser than 0.8 interesting degree are the low interesting patterns, the patterns whose interesting degrees are higher than 1 interesting degree are the high interesting patterns, and the patterns whose interesting degree between 0.8 and 1 are the common patterns. So $p_6 = \{u_1, d_2, l_3, t_{17}, s_3\}$ and $p_{10} = \{u_2, d_1, l_3, t_{17}, s_3\}$ are the low interesting degree patterns, $p_9 = \{u_2, d_2, l_3, t_{17}, s_3\}$, $p_{14} = \{u_3, d_2, l_5, t_{18}, s_5\}$, and $p_{17} = \{u_3, d_1, l_5, t_{18}, s_3\}$ are the common interesting degree patterns, and $p_5 = \{u_1, d_1, l_3, t_{17}, s_3\}$, $p_{11} = \{u_2, d_2, l_2, t_{17}, s_5\}$, and $p_{16} = \{u_3, d_1, l_5, t_{18}, s_3\}$ are the high interesting degree patterns. We delete the low interesting degree patterns and get the patterns with high interesting degree as is shown in Table 5.

Then, we use the Apriori algorithm to mine rules on the high interesting degree patterns; we set the minimum support to 20% and the minimum confidence to 80%, then we can get the results as follows.

The lift denotes the ratio of the confidence to the support of the consequent item; the computational formula is followed: $L_{x \rightarrow y} = C_{x \rightarrow y} / S_y$. The lift reacts the influence degree of the antecedent item X to the consequent item Y appears. Generally, the lift value should be larger than 1, and it means that the antecedent item X has a positive influence on the consequent item Y appears. The larger the life value is, the better the rule is.

From Table 6, we can conclude that we can get 39 association rules when we use the method we proposed in this paper. These rules were obtained from the high interesting pattern; we thought that these rules were interesting rules. Then we observe the rule with the maximum lift, time = t_{17} , and service = $s_3 \rightarrow$ location = l_3 . The value is 4.5. It means that this association rule has the highest realistic guidance. So this rule will be firstly considered when we use the rules of the result. We can use these association rules to recommend some services to tourist to offer them better services; for example, using the rule location = l_2 and time = $t_{16} \rightarrow$ service = s_4 , we can recommend the s_4 to the tourist when the tourist stays in the context with location = l_2 and time = t_{16} . In this paper, the service s_4 is the tourism route guide, so we can send the tourism route guide to the tourist as is shown in Figure 11.

4.2. Comparison and Discussion. To verify the effects of the method we proposed in this paper, we use the Apriori

FIGURE 7: The network we constructed.

FIGURE 8: Calculate the tourist u_1 collection weight. FIGURE 9: Calculate the tourist u_2 collection weight.

FIGURE 10: Calculate the tourist u_3 collection weight.

FIGURE 11: Tourism route guide.

algorithm, the GRI algorithm, the CARAMA algorithm and Predictive-Apriori algorithm on the original data (here we set the minimum support equals to 20% and the minimum confidence equals to 80%; too), and we get following rules as is shown in Tables 7, 8, 9, and 10.

Comparing Table 6 with Tables 7 and 8, there are 11 rules from Table 7 which have been emerged in Table 6 (the rules marked with yellow as is shown in Table 6), and all rules in Table 8 have been emerged in Table 6. So we think that the method we proposed to mine the mobile customer behavior pattern has the merit of effectiveness; in this experiment the validity of the method is about 91.67% (11/12) to the Apriori algorithm and 100% (6/6) to the GRI algorithm, which means the coverage values are 91.67% and 100%, which are larger than the threshold we set before. It means that the method we proposed is feasible and effective. Excluding the 11 rules

in Table 7, Table 6 has other 28 rules and these rules have the feature of high interest, so they will provide more choices to the service provider and more services to the mobile customer. Then we observe the rule which has the maximum value of lift from Tables 6 and 7, the rule is $\text{time} = t_{17}$ and $\text{service} = s_3 \rightarrow \text{location} = l_3$; it means that the method we proposed is similar to the classical Apriori algorithm. At last, the rule whose ID = 1 in Table 7: $\text{location} = l_1 \rightarrow \text{user} = u_1$. It is the only rule that is not included in Table 6, although this rule meets the minimum support and the minimum confidence; the pattern with $\{l_1, u_1\}$ is a low interesting pattern as we definite before, and the rule $\text{location} = l_1 \rightarrow \text{user} = u_1$ is an uninteresting rule. In our method, we can reject uninteresting rules like this. Through the analysis, the method we proposed in this paper is more feasible and advanced when being compared with the Apriori algorithm.

TABLE 6: Mining results.

ID	Rules	Count	Lift
1	User = u_1 $\rightarrow$ device = d_1	2	3
2	Location = l_3 $\rightarrow$ time = t_{17}	2	3
3	Location = l_3 $\rightarrow$ service = s_3	2	2.25
4	Service = s_5 $\rightarrow$ device = d_2	2	1.5
5	User = u_2 $\rightarrow$ device = d_2	3	1.5
6	Time = t_{16} $\rightarrow$ service = s_4	3	3
7	Service = s_4 $\rightarrow$ time = t_{16}	3	3
8	Time = t_{18} $\rightarrow$ user = u_3	3	2.25
9	Time = t_{18} $\rightarrow$ location = l_5	3	2.25
10	User = u_3 $\rightarrow$ location = l_5	4	2.25
11	Location = l_5 $\rightarrow$ user = u_3	4	2.25
12	Location = l_3 and time = t_{17} $\rightarrow$ service = s_3	2	2.25
13	Location = l_3 and service = s_3 $\rightarrow$ time = t_{17}	2	3
14	Time = t_{17} and service = s_3 $\rightarrow$ location = l_3	2	4.5
15	User = u_2 and location = l_2 $\rightarrow$ device = d_2	2	1.5
16	Location = l_2 and device = d_2 $\rightarrow$ user = u_2	2	3
17	User = u_2 and time = t_{17} $\rightarrow$ device = d_2	2	1.5
18	Time = t_{17} and device = d_2 $\rightarrow$ user = u_2	2	3
19	Location = l_2 and time = t_{16} $\rightarrow$ service = s_4	2	3
20	Location = l_2 and service = s_4 $\rightarrow$ time = t_{16}	2	3
21	Time = t_{16} and device = d_2 $\rightarrow$ service = s_4	2	3
22	Service = s_4 and device = d_2 $\rightarrow$ time = t_{16}	2	3
23	Time = t_{18} and user = u_3 $\rightarrow$ location = l_5	3	2.25
24	Time = t_{18} and location = l_5 $\rightarrow$ user = u_3	3	2.25
25	Time = t_{18} and service = s_3 $\rightarrow$ user = u_3	2	2.25
26	User = u_3 and service = s_3 $\rightarrow$ time = t_{18}	2	3
27	Time = t_{18} and device = d_2 $\rightarrow$ user = u_3	2	2.25
28	Time = t_{18} and service = s_3 $\rightarrow$ location = l_5	2	2.25
29	Location = l_5 and service = s_3 $\rightarrow$ time = t_{18}	2	3
30	Time = t_{18} and device = d_2 $\rightarrow$ location = l_5	2	2.25
31	User = u_3 and service = s_3 $\rightarrow$ location = l_5	2	2.25
32	Location = l_5 and service = s_3 $\rightarrow$ user = u_3	2	2.25
33	User = u_3 and device = d_2 $\rightarrow$ location = l_5	3	2.25
34	Location = l_5 and device = d_2 $\rightarrow$ user = u_3	3	2.25
35	Time = t_{18} , user = u_3 and service = s_3 $\rightarrow$ location = l_5	2	2.25
36	Time = t_{18} , location = l_5 and service = s_3 $\rightarrow$ user = u_3	2	2.25
37	User = u_3 , location = l_5 and service = s_3 $\rightarrow$ time = t_{18}	2	3
38	Time = t_{18} , user = u_3 and device = d_2 $\rightarrow$ location = l_5	2	2.25
39	Time = t_{18} , location = l_5 and device = d_2 $\rightarrow$ user = u_3	2	2.25

5. Conclusion

In this paper we considered the context factors which influence the tourist behavior pattern comprehensively, such as the device the tourist use, time, location, and service types, and got the context set which influences the tourist behavior pattern. Then we proposed a method to mine tourist behavior patterns based on the network diagram; this method constructed a network diagram firstly. Then, we got the behavior patterns with high interesting degree and did association rule mining in the patterns and got the rules;

at last, we made an experiment to show the feasibility and effectiveness of our method. In our experiment, we set the low interest degree th_1 value to be equal to 0.8 and the high interest degree th_2 value to be equal to 1 and deleted the low interest pattern; then we did association mining with Apriori algorithm to the remainder of the patterns and got 39 rules; we can do some recommendations to the tourist with these high interest rules. Compared to the results which do not use this method, it has the following advantages: (1) it can keep the interest rules and delete the uninterested rules in the results; (2) it can produce many other interest rules, which

TABLE 7: The results based on Apriori algorithm.

ID	Rules	Count	Lift
1	Location = l_1 $\rightarrow$ user = u_1	4	2.57
2	Time = t_{18} $\rightarrow$ user = u_3	4	3.00
3	Time = t_{18} $\rightarrow$ location = l_5	4	3.00
4	Location = l_3 $\rightarrow$ time = t_{17}	4	3.00
5	Location = l_3 $\rightarrow$ service = s_3	4	2.57
6	User = u_3 $\rightarrow$ location = l_5	6	3.00
7	Location = l_5 $\rightarrow$ user = u_3	6	3.00
8	Time = t_{18} and user = u_3 $\rightarrow$ location = l_5	4	3.00
9	Time = t_{18} and location = l_5 $\rightarrow$ user = u_3	4	3.00
10	Location = l_3 and time = t_{17} $\rightarrow$ service = s_3	4	2.57
11	Location = l_3 and service = s_3 $\rightarrow$ time = t_{17}	4	3.00
12	Time = t_{17} and service = s_3 $\rightarrow$ location = l_3	4	4.50

TABLE 8: The results based on GRI algorithm.

ID	Rules	Count	Lift
1	Location = l_1 $\rightarrow$ user = u_1	4	2.5
2	Location = l_3 $\rightarrow$ service = s_3	4	2.5
3	User = u_3 $\rightarrow$ location = u_3	6	3
4	Location = l_5 $\rightarrow$ user = u_3	6	3
5	Location = l_3 $\rightarrow$ time = t_{17}	4	3
6	Time = t_{17} and service = s_3 $\rightarrow$ location = l_3	4	4.5

TABLE 9: The results based on CARMA algorithm.

ID	Rules	Count	Lift
1	User = u_3 $\rightarrow$ location = l_5	6	3
2	Location = l_5 $\rightarrow$ user = u_3	6	3
3	Location = l_1 $\rightarrow$ user = u_1	4	2.5
4	Location = l_3 $\rightarrow$ time = t_{17}	4	3
5	Location = l_3 $\rightarrow$ service = s_3	4	2.5
6	Location = l_3 $\rightarrow$ time = t_{17} and service = s_3	4	4.5
7	Time = t_{18} $\rightarrow$ user = u_3	4	3
8	Time = t_{18} $\rightarrow$ location = l_5	4	3
9	Location = l_3 and time = t_{17} $\rightarrow$ service = s_3	4	2.5
10	Location = l_3 and service = s_3 $\rightarrow$ time = t_{17}	4	3
11	Time = t_{17} and service = s_3 $\rightarrow$ location = l_3	4	4.5

we can use them to make more recommendations for the tourist; (3) it can produce the same rule which has the highest lift compared to the result that does not use this method. That is, the method we used in this paper is feasible and superior.

The future work will be further researching on the context factors which influence the tourist behavior pattern and expanding the context set; also we will analyze the performance of the method we proposed and optimize the method and so on.

TABLE 10: The results based on Predictive-Apriori algorithm.

ID	Rules
1	User = u_3 $\rightarrow$ location = l_5
2	Location = l_5 $\rightarrow$ user = u_3
3	Location = l_1 $\rightarrow$ user = u_1
4	Location = l_3 $\rightarrow$ time = t_{17} and service = s_3
5	Time = t_{18} $\rightarrow$ user = u_3 and location = l_5
6	Time = t_{17} and service = s_3 $\rightarrow$ location = l_3
7	Service = s_2 $\rightarrow$ user = u_1 and location = l_1
8	Time = t_{15} $\rightarrow$ user = u_1
9	User = u_1 and service = s_3 $\rightarrow$ location = l_3
10	User = u_2 and device = d_1 $\rightarrow$ service = s_3
11	User = u_3 and service = s_3 $\rightarrow$ location = l_5 and time = t_{18}
12	Location = l_2 and time = t_{16} $\rightarrow$ service = s_4
13	Location = l_2 and service = s_4 $\rightarrow$ time = t_{16}
14	Location = l_5 and service = s_3 $\rightarrow$ user = u_3 time = t_{18}
15	User = u_1 and device = d_2 and location = l_1 $\rightarrow$ service = s_2

Appendix

Questionnaire

Your age: —

Gender: male/female

(1) To what extent do you think the user will influence the behavior?

- (A) Strongly disagree
- (B) Disagree
- (C) Neutral
- (D) Agree
- (E) Strongly agree

(2) To what extent do you think the location will influence the behavior?

- (A) Strongly disagree
- (B) Disagree
- (C) Neutral
- (D) Agree
- (E) Strongly agree

(3) To what extent do you think the time will influence the behavior?

- (A) Strongly disagree
- (B) Disagree
- (C) Neutral
- (D) Agree
- (E) Strongly agree

- (4) To what extent do you think the device will influence the behavior?
- (A) Strongly disagree
 - (B) Disagree
 - (C) Neutral
 - (D) Agree
 - (E) Strongly agree
- (5) To what extent do you think the service will influence the behavior?
- (A) Strongly disagree
 - (B) Disagree
 - (C) Neutral
 - (D) Agree
 - (E) Strongly agree
- (6) To what extent do you think the network bandwidth will influence the behavior?
- (A) Strongly disagree
 - (B) Disagree
 - (C) Neutral
 - (D) Agree
 - (E) Strongly agree
- (7) To what extent do you think the temperature will influence the behavior?
- (A) Strongly disagree
 - (B) Disagree
 - (C) Neutral
 - (D) Agree
 - (E) Strongly agree
- (8) To what extent do you think the law will influence the behavior?
- (A) Strongly disagree
 - (B) Disagree
 - (C) Neutral
 - (D) Agree
 - (E) Strongly agree
- (9) Others factors will influence the behavior, such as —.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant nos. 71071140 and 71301070005), the National Natural Science Foundation of Zhejiang Province (Grant no. Y1090617), the Key Innovation Team of Zhejiang Province (Grant no. 2010R50041), the Soft science key research project of Zhejiang Province (Grant no. 2013C25053), the Zhejiang Gongshang University Graduate Student Scientific Research Project (1130XJ1512168), and the Modern Business Centre of Zhejiang GongShang University.

References

- [1] M. Qing, *Research on Constructing Personalized Tourism Service Supply Chain in the Context of Modern Information Technology*, Shandong University of Finance and Economics, 2012.
- [2] A. Farmaki, "An exploration of tourist motivation in rural settings: the case of Troodos, Cyprus," *Tourism Management Perspectives*, vol. 2-3, pp. 72–78, 2012.
- [3] C. A. Martin and S. F. Witt, "Tourism demand forecasting models: choice of appropriate variable to represent tourists' cost of living," *Tourism Management*, vol. 8, no. 3, pp. 233–246, 1987.
- [4] C. Smallman and K. Moore, "Process studies of tourists' decision-making," *Annals of Tourism Research*, vol. 37, no. 2, pp. 397–422, 2010.
- [5] S. S. Kim, D. J. Timothy, and J. Hwang, "Understanding Japanese tourists' shopping preferences using the Decision Tree Analysis method," *Tourism Management*, vol. 32, no. 3, pp. 544–554, 2011.
- [6] B. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in *Proceedings of the Workshop on Mobile Computing Systems and Applications*, pp. 85–90, December 1994.
- [7] P. J. Brown, J. D. Bovey, and X. Chen, "Context-aware applications: from the laboratory to the marketplace," *IEEE Personal Communications*, vol. 4, no. 5, pp. 58–64, 1997.
- [8] Y. A. Rakotonirain, S. W. Iokey, and G. Fitzpatrick, "Context-awareness for the mobile environment," 2008-10-02.
- [9] D. Snowdon and A. Grasso, "Providing context awareness via a large screen display," in *Proceedings of the CHI 2000 Workshop on "The What, Who, Where, When, Why and How of Context-Awareness"*, 2000, report no. 2000/011.
- [10] J. Z. Gu, "Context aware computing," *Journal of East China Normal University: Natural Science*, vol. 5, pp. 1–20, 2009.
- [11] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207–216, May 1993.
- [12] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, China Machine Press, Beijing, China, 3rd edition, 2012.
- [13] K. J. Kang, B. Ka, and S. J. Kim, "A service scenario generation scheme based on association rule mining for elderly surveillance system in a smart home environment," *Engineering Applications of Artificial Intelligence*, vol. 25, pp. 1355–1364, 2012.
- [14] Z. Y. Zhang, Y. P. Zhang, J. Y. Zhang, and X. J. Zhang, "Application of improved association rule algorithm in college teaching management," *Computer Engineering*, vol. 38, no. 2, pp. 75–78, 2012.
- [15] V. S. Tseng and K. W. Lin, "Efficient mining and prediction of user behavior patterns in mobile web systems," *Information and Software Technology*, vol. 48, no. 6, pp. 357–369, 2006.
- [16] Z. B. Ma, T. J. Lu, and H. Q. Li, "Mining of temporal sequence mobile access patterns based on context awareness," *Transaction of Beijing Institute of Technology*, vol. 28, no. 10, pp. 937–940, 2008.
- [17] T. S. Chen, Y. S. Chou, and T. C. Chen, "Mining user movement behavior patterns in a mobile service environment," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 42, no. 1, pp. 87–101, 2012.
- [18] Y. Q. Hu, *Operational Research*, Tsinghua University Press, Beijing, China, 3rd edition, 2007.