Cluster-Based Temporal Mobile Sequential Patterns Prediction Strategy Using Lbs

Manochitra.V
Assistant Professor
Rajalakshmi Institute of Technology, Chennai

Abstract: The advancement in Wireless communication techniques with mobile phones and GPS enabled cellular phones have introduced to a new business model. This emerges to a new way of service called as Location-Based Service (LBS).It is defined as any service that takes into account the geographic location of an entity in order to deliver a service. Mining user behaviour in mobile environments is an emerging and important task in data mining. Existing impact focus on mobile patterns from logs and their impact can't deliver the predictions since the constant change in the mobile behaviour .In this project a new data mining algorithm called Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine), to identify the patterns and the strategy to predict the behaviour of the mobile user is proposed. Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) is used to identify the user clusters .Location-Based Service alignment helps in finding the similarities between users. An approach for Time segmentation is provided to find the time intervals where similar mobile characteristics exist. Here the concepts of mining and prediction of mobile behaviours with user relations and temporal property with a mutual understanding deliberately exhibit the data explicitly with the User Interface. The prediction strategy uses mobile patterns to predict the mobile user behaviour in near future. Considering user clusters and time segmentation simultaneously complete information concerning personal mobile behaviour is predicted.

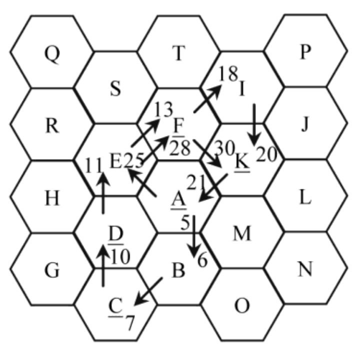
Keywords: Location-Based Service(LBS), Cluster-based Temporal Mobile Sequential Pattern Mine.

I. INTRODUCTION

Researches on Location-Based Service (LBS) have been emerging in recent years due to a wide range of potential applications. One of the active topics is the mining and prediction of mobile movements and associated transactions. Most of existing studies focus on discovering mobile patterns from the whole logs. However, this kind of patterns may not be precise enough for predictions since the differentiated mobile behaviors among users and temporal periods are not considered. In this paper, we propose a novel algorithm, namely, Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine), to discover the Cluster-based Temporal Mobile Sequential Patterns (CTMSPs). Moreover, a prediction strategy is proposed to predict the subsequent mobile behaviors. In CTMSP-Mine, user clusters are constructed by a novel algorithm named Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) and similarities between users are evaluated by the proposed measure, Location-Based Service Alignment (LBS-Alignment). Meanwhile, a time segmentation approach is presented to find segmenting time intervals where similar mobile characteristics exist. HE advancement of wireless communication techniques and the popularity of mobile devices such as mobile phones, PDA, and GPS-enabled cellular phones, have contributed to a new business model. Mobile users can request services through their mobile devices via Informa- tion Service and Application Provider (ISAP) from any- where at any time . This business model is known as Mobile Commerce (MC) that provides Location-Based Services (LBS) through mobile phones. MC is expected to be as popular as e-commerce in the and it is based on the cellular network composed of several base stations. The communication coverage of each base station is called a cell as a location area. The average distance between two base stations is hundreds of meters and the number of base stations is usually more than 10,000 in a city. When users move within the mobile network, their locations and service requests are stored in a centralized mobile transaction database. Fig. 1 shows an MC scenario, where a user moves in the mobile network and requests services in the corresponding cell through the mobile devices. A moving sequence of a user, where cells are underlined if services are requested there. In the record of service transactions, where the service S_1 was requested when this user moved to the location A at time 5. In fact, there exists insightful information in these data, such as movement and transaction behaviors of mobile users. Mining mobile transaction data can provide insights for various applications, such as data prefetching and service recommendations. A mobile transaction database is complicated since a huge amount of mobile transaction logs is produced based on the user's mobile behaviors. Data mining is a widely used technique for discovering valuable information in a complex data set and a number of studies have discussed the issue of mobile behavior mining. The main difference between these literatures is the involved information of proposed patterns. Tseng and Tsui addressed the problem of mining associated service patterns in mobile web networks. Tseng and Lin also proposed SMAP-Mine to efficiently mine users' sequential mobile access patterns, based on the FP-Tree. Chen et al. proposed the path traversal patterns for mining mobile web user behaviors. Yun and Chen proposed a novel method of mining mobile sequential patterns. To increase the accuracy of predictions, the moving path was taken into consideration in the above studies. However, mobile behaviors vary among different user clusters or at various time intervals. The prediction of mobile behavior will be more precise if we can find the corresponding mobile patterns in each user cluster and time interval. To provide precise location-based services for users, effective mobile behavior mining systems are required pressingly. Clustering mobile transaction data helps in the discovery of social groups, which are used in applications such as targeted advertising, shared data allocation, and personali- zation of content services. In previous studies, users are typically clustered according to their personal profiles (e.g., age, sex, and occupation). However, in real applications of mobile environments, it is often difficult to obtain users' profiles. That is, we may only have access to users' mobile transaction data. To achieve the goal of user clustering without user profiles, we need to evaluate the similarities of mobile transaction sequences (MTSs). Although a number of clustering algorithms have been studied in the rich literature, they are not applicable in the LBS scenario in consideration of the following issues: 1) Most clustering methods in can only process data with spatial similarity measures, while clustering methods with nonspatial similarity measures are required for LBS environments. 2) Most clustering methods in request the users to set up some parameters. However, in real applications, it is difficult to determine the right parameters manually for the clustering tasks. Hence, an automated clustering method is required. Although there exist many nonspatial similarity measures , most of them are used to measure the string similarity. However, the mobile transaction sequences discussed in this paper include multiple and heterogeneous information such as time, location, and services. Therefore, the existing measures are not applicable directly for measuring the similarity of mobile transaction sequences. To our best knowledge, this is the first work on mining and prediction of mobile behaviors with considerations of user elations and temporal property simultaneously. Through experimental evaluation under various simulated conditions, the proposed methods are shown to deliver excellent performance. Here we have classified the section as follows. The section 2 deals about the existing work in the prediction strategies. The section 3 deals about the proposed method. The section 4 deals about the results & Discussions. The conclusion is given at the final section.

II. EXISTING METHODS

In an Existing System, we don't use Cluster-based Temporal Mobile Sequential Patterns (CTMSPs). Since Previous studies and applications consider time to be an important factor. Users have some specific behaviors in specific time and also data's not maintained in a centralized way. A mobile transaction database is complicated since a huge amount of mobile transaction logs is produced based on the user's mobile behaviors. Data mining is a widely used technique for discovering valuable information in a complex data set and a number of studies have discussed the issue of mobile behavior mining. The main difference between these literatures is the involved information of proposed patterns.



Time	Cell	Service
5	A	S_1
6	B	
7	C	S_3
10	D	S_2
11	E	
13	F	S_3S_4
18	I	
20	K	S_5
21	A	S_1
25	E	
28	F	
30	K	S_2

(a) (b)

Fig. 1. An example for a mobile transaction sequence. (a) Moving sequences. (b) Service sequences.

The time interval segmentation method helps us find various user behaviors in different time intervals. For example, users may request different services at different times (e.g., day or night) even in the same location. If the time interval factor is

International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.1, No.6, October 2012 DOI:10.15693/ijaist/2012.v1i6.7-13

not taken into account, some behaviors may be missed during specific time intervals. To find complete mobile behavior patterns, a time interval table is required. Although some studies used a predefined time interval table to mine mobile patterns [11], [19], the data characteristic and data distribution vary in real mobile applications. Therefore, it is difficult to predefine a suitable interval table by users. Automatic time segmentation methods are, thus, required to segment the time dimension in a mobile transaction database. a novel data mining algorithm named Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine) to efficiently mine the Cluster-based Temporal Mobile Sequential Patterns (CTMSPs) of users. Then, novel prediction strategies are proposed to effectively predict the user's subsequent behaviors using the discovered CTMSPs. To mine CTMSPs, we first propose a transaction clustering algorithm named Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) that builds a cluster model for mobile transactions based on the proposed Location-Based Service Alignment (LBS-Alignment) similarity measure. Then, we take advantage of the Genetic Algorithm (GA) to produce a more suitable time interval table. Based on the produced user clusters and the time interval table, all CTMSPs can be discovered by the proposed method. To our best knowledge, this is the first work on mining and prediction of mobile sequential patterns by considering user clusers and temporal relations in LBS environments simultaneously. Finally, through experimental evaluation on various simulated conditions, the proposed method is shown to deliver excellent performance in terms of precision, recall, and F-measure.

The main contributions of this work are that we propose not only a novel algorithm for mining CTMSPs but also two nonparametric techniques for increasing the predictive precision of the mobile users' behaviors. Besides, the proposed CTMSPs provide information including both user clusters and temporal relations. Meanwhile, user profiles like personal information are not needed for the clustering method and time segmentation method

III. PROPOSED SYSTEM

This Proposed system we implement new technique called CTMSPs. To mine CTMSPs, we first propose a transaction clustering algorithm named Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) that builds a cluster model for mobile transactions based on the proposed Location-Based Service Alignment (LBS-Alignment) similarity measure.

Then, we take advantage of the Genetic Algorithm (GA) to produce a more suitable time interval table. Based on the produced user clusters and the time interval table, all CTMSPs can be discovered by the proposed method.

To our best knowledge, this is the first work on mining and prediction of mobile sequential patterns by considering user clusters and temporal relations in LBS environments simultaneously. The main contributions of this work are that we propose not only a novel algorithm for mining CTMSPs but also two nonparametric techniques for increasing the predictive precision of the mobile users' behaviors.

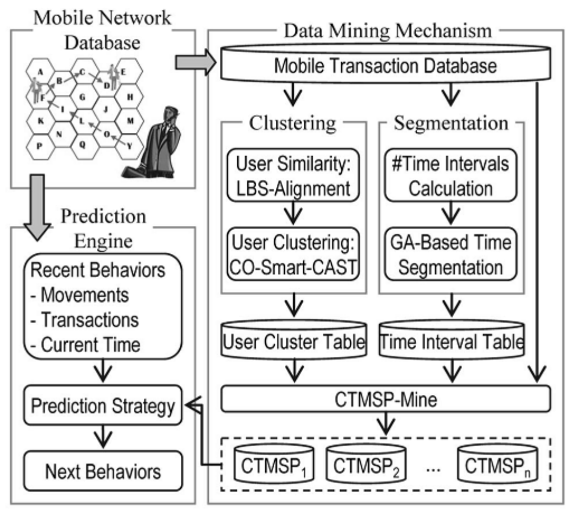


Fig. 3. System Framework

In this section, we describe our system design. Four important research issues are addressed here:

1. Clustering of mobile transaction sequences.
2. Time segmentation of mobile transaction sequences.
3. Discovery of CTMSPs.
4. Mobile behavior prediction for mobile users.

Fig. 2 shows the proposed system framework. Our system has an "offline" mechanism for CTMSPs mining and an "online" engine formobile behavior prediction.Whenmobile users move within the mobile network, the information which includes time, locations, and service requests will be stored in the mobile transaction database. Table 1 shows an example ofmobile transaction databasewhich contains seven records. In the offline dataminingmechanism,we design two techniques and the CTMSP-Mine algorithm to discover the knowledge. First,we propose theCO-Smart-CAST algorithm to cluster themobile transaction sequences. In this algorithm, we propose the LBS-Alignment to evaluate the similarity of mobile transaction sequences. Second, we propose a GA- based time segmentation algorithm to find the most suitable time intervals. After clustering and segmentation, a user

```
Input: Two mobile transaction sequences s and s'
Output: The similarity between s and s'
LBS-Alignment (s, s')
  p ← 0.5 / (s.length + s'.length) /*p is the location penalty*/
  M_{0,0} ← 0.5
  M_{i,0} ← M_{i-1,0} – p, ∀ i = {1, 2, ..., s.length}
  M_{0,j} ← M_{0,j-1} – p, ∀ j = {1, 2, ..., s'.length}
  For i ← 1 to s.length
    For j ← 1 to s'.length
      If s_i.location = s_j'.location
        TP ← p × | s_i.time - s_j'.time| / len /*time penalty*/
        SR ← p × (s_i.service ∩ s_j'.service) / (s_i.service ∪ s_j'.service) /*service reward*/
        M_{i,j} ← Max(M_{i-1,j-1} – TP + SR, M_{i-1,j} - p, M_{i,j-1} - p)
      Else
        M_{i,j} ← Max(M_{i-1,j} - p, M_{i,j-1} - p)
      End If
    End For
  End For
  Return M_{s.length,s'.length}
```

Table1: LDS Algorithm

cluster table and a time interval table are generated,respectively. Third, we propose the CTMSP-Mine algorithm to mine the CTMSPs from the mobile transaction database according to the user cluster table and the time interval table. In the online prediction engine, we propose a behavior prediction strategy to predict the subsequent behaviors according to the mobile user's previous mobile transaction sequences and current time. The main purpose of this framework is to provide mobile users a precise and efficient mobile behavior prediction system. In a mobile transaction database, users in the different user groups may have different mobile transaction behaviors. The first task we have to tackle is to cluster mobile transaction sequences. We proposed a parameter-less clustering algorithm CO-Smart-CAST. Before performing the CO-Smart-CAST, we have to generate a similarity matrix S, based on the mobile transaction database. The entry Sij in matrix S represents the similarity of the mobile transaction sequences i and j in the database, with the degrees in the range of [0, 1]. A mobile transaction sequence can be viewed as a sequence string, where each element in the string indicates a mobile transaction. The major challenge we have to tackle is to measure the content similarity between mobile transactions. We propose LBS-Alignment, which can obtain the similarity based on the concept of DNA alignment [4]. LBS-Alignment is based on the consideration that two mobile transaction sequences aremore similar, when the orders and timestamps of their mobile transactions are more similar. Based on this concept,we specifically design the time penalty (TP) and the service reward (SR) in the LBS-Alignment. The base similarity score is set as 0.5. Twomobile transactions can be aligned if their locations are the same. Otherwise, a location penalty is generated to decrease their similarity score. The location penalty is defined as 0:5=ðjs1jþjs2jÞ, where js1j and js2j are the lengths of sequences s1 and s2, respectively. Notice that the maximal number of location penalties is js1jþjs2j. When two sequences are totally different, their similarity score is 0. When two mobile transactions are aligned, we measure their time penalty and service reward. TP focuses on their time distance. The farther the time distances between them, the larger their time pesnalty. TP that is generated to decrease their similarity score is defined as (ðjs1 time s2 timejÞ=len,where len indicates the time length. SR focuses on the similarity of the service requests. The more similar their service requests, the larger their service reward. SR that is generated to increase their similarity score is defined as js1:services \ s2:servicesj=js1:services [s2:servicesj.

Segmentation of Mobile Transactions

In a mobile transaction database, similar mobile behaviors exist under some certain time segments. Hence, it is important to make suitable settings for time segmentation so as to discriminate the characteristics of mobile behaviors under different time segments. We propose a GA-based method to automatically obtain the most suitable time segmentation table with common mobile behaviors. shows the procedure of our proposed time segmentation method, named Get Number of Time Segmenting Points (GetNTSP) algorithm. The input data are a mobile transaction database D and its time length T (line 01). The output data are the number of time segmenting points (line 02). For each item, we accumulate the total number of occurrences at each time point (line 07 to line 11). Therefore, an item (location, service) can draw a curve of count distribution, as shown in Fig. 8. For all curves, we found the time points with the largest change rate (line 13).We defined the change rate as (c½i þ 1 c½i Þ=ð1þc½i), where c½i represents the total number of occurrences for the item at time point i.We count occurrences of all these time points (line 15), and find out the satisfied time points whose counts are larger than or equal to the average of all occurrences from these ones, and then, take these satisfied ones as a set of the time point sequence (TPS) (line 17). In the time point sequence, we calculate the average time distance a between two neighboring time points (line 18). We calculate the number of neighboring time point pairs, in which the time distance is higher than a (line 19 to line 23). The result represents the time segmentation count (line 24). 12 pairs of locations and services, and their time points with the largest change rates are 5, 10, 13, 30, 5, 28, 10, 7, 20, 30, 25, and 28. These time points can be sorted as 5(2), 7(1), 10(2), 13(1), 20(1), 25(1), 28(2), and 30(2), where t(n) indicates that the number of time points t is n. The TPS is {5, 10, 28, and 30} because the average number of time points is 12=8 ¼ 1:5. We calculate that a is 25=3 ¼ 8:33. There is one interval between 10 and 28 which is larger than a in TPS. Therefore, the number of time segmenting points is obtained to be 1. After we obtain the number of time segmenting points, we use the genetic algorithm to discover the most suitable time intervals. In GA, a chromosome with a length equal to the number of time segmenting points is defined as a time segmenting point set. For instance, let C ¼f30; 60; 120g be the representation of chromosome. C includes three time segmenting points to segment the mobile transactions into four time intervals. Initially, we randomly generate the initial population and define a suitable fitness function. Through repeated selection, crossover, and mutation, we obtain an optimal solution. There are three operators in Genetic Algorithm: 1) selection, 2) crossover, and 3) mutation. For the selection operator, a proportion of the current population is selected to product the next population in each generation. Individual chromosomes are selected based on their fitness value. The larger the fitness value of a chromosome, the higher the probability of the chromosome is selected. For the crossover operator, we apply one-point crossover that involves a crossover probability to this operator. A crossover point on both parent chromosomes is randomly selected. All time segmenting points beyond the crossover point are swapped between the two parent chromosomes. The resulting chromosomes are the children. For the mutation operator, we apply the one-bit mutation to this operator. This operator involves a mutation probability that arbitrary time segmenting point in a chromosome will be changed from its original state. For any children chromosome, its time segmenting points must be sorted if the orders are not progressively increased.

IV. RESULTS & DISCUSSIONS

In this section, we conducted a series of experiments to evaluate the performance of the proposed CTMSP-Mine, under various system conditions. Experiments can be divided into three parts: 1) user clustering, 2) time interval segmentation, and 3) precision of prediction. All of the experiments were implemented in Java on a 3.0 GHz machine with 1 GB of memory running Windows XP.

4.1 Simulation Model

To evaluate the practicability of our prediction strategies, we simulate the conditions of the real world to generate the required network environment and transaction data set. The simulation model is referred by Yun and Chen [37]. Table 6 shows the major parameters with the default setting. In the base experiment model, the network is modeled as a jWj jWj mesh network [21], where each node represents one cell [26]. The atomic temporal observation unit of life is "day." There are Ntp time points in a day. We simulate NT time intervals and the start time points and end time points are determined from a normal distribution. In each interval, the number of services in each cell is determined from a uniform distribution within a given range Ns. For each cell, the advancing probability Pa of each neighbor is the probability for a user to move to neighboring cells and request the services, i.e., each directed edge from one cell to another cell is assigned with an advancing probability which is defined as the ratio of the number of services in each neighbor to those numbers Of other neighbors. The backward moving represents that a user will move from the current cell back to the cell from which he came. The backward probability is denoted by P0 ¼ Pa Wb, whereWb is a backward weight. For example, Fig. a shows a 3X 3 mesh network. There are four

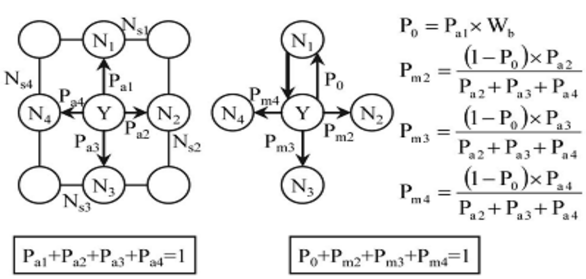


Fig. 4. Simulate an LBS environment using mesh network. (a) 3X3 mesh network. (b) When coming from node N1.

neighbors N1;N2;N3, and N4 for cell Y and the corresponding advancing probabilities are Pa1;Pa2;Pa3,and Pa4, respectively. In addition, Ns1;Ns2;Ns3, and Ns4 indicate the numbers of services in cells N1;N2;N3, and N4 and Pai ¼ Nsi=ðNs1 þ Ns2 þ Ns3 þ Ns4Þ; 81 i 4. When a user visits cell Y from cell N1, the probabilities of the neighbors are shown in Fig. In this experiment, there are N users in this network. We divide all users into NU user groups equally. When a user moves within cells for requesting in the network, the mobile transaction sequence is generated by this user. The method for generating mobile transaction sequences is referred to [23]. The length of each mobile transaction sequence is determined from a Poisson distribution with mean equal to LT . When a user moves to a cell, the probability that this user requests the service in this cell is denoted by Psr.In [37], the paper mentioned that people tend to buy sets of items together, which are celled potential maximal frequent sets. A transaction may contain one or more of such frequent sets. Therefore, in each user group, some potential events are generated to represent the mobile behaviors in this user group. The potential events indicate the preferred service sequences which users frequently request for each user group. Each user may request services by adhering to potential events with probability PE or randomly [30], [31]. The followings are the main measurements for the experimental evaluation. The precision, recall, and F measure are defined as (5), (6), and (7), where pþ and p indicate the number of correct predictions and incorrect predictions, respectively, and jRj indicates the total number of service requests.

$$Precision = \frac{p^+}{p^+ + p^-},$$
$$Recall = \frac{p^+ + p^-}{|R|},$$
$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}.$$

4.2 Comparison of Various Clustering methods

This experiment analyzes the effects of LBS-Alignment and CO-Smart-CAST, when the number of predefined user groups is varied. First, five data sets with various number of user groups are generated. Fig. 13a shows that LBS- Alignment outperforms LCSS, DTW, EDR, and ERP in terms of classification accuracy. The main reason is that LBS- Alignment takes into account the similarities of time series string and categorical data at the same time, while LCSS, DTW, EDR, and ERP consider only the similarity of time series strings.

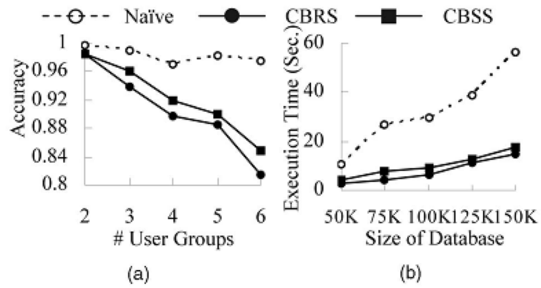


Fig. 5. Results of various KNN strategies. (a) Accuracy. (b) Execution time.

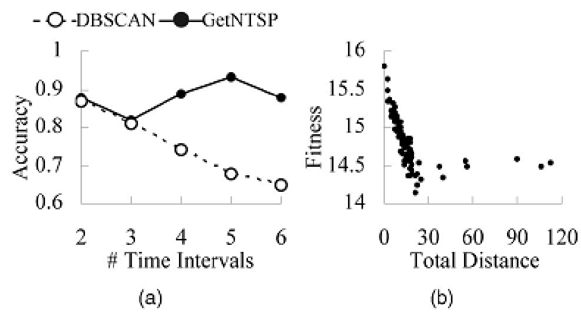


Fig.6. Results of time segmentation algorithms. (a) Accuracy. (b) Fitness.

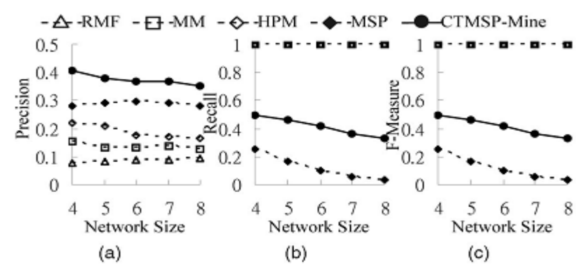


Fig. 7. Results with network size varied. (a) Precision. (b) Recall (c) F-Measure.

A series of experiments were conducted for evaluating the performance of the proposed methods. The experimental results show that CO-Smart-CAST method achieves high-quality clustering results and the proposed CBSS strategy obtains highly precise results for user classification. Meanwhile, our GA-based method obtains the most proper and correct time intervals. For behavior prediction, CTMSP is shown to outperform other prediction methods in terms of precision and F-measure. The experimental results demonstrate that our proposed methods are efficient and accurate under various conditions.

V. CONCLUSION

In this paper, we have proposed a cluster-based temporal mobile sequential patterns prediction strategy using lbs, for discovering CTMSPs in LBS environments. Furthermore, we have proposed novel prediction strategies to predict the subsequent user mobile behaviors using the discovered CTMSPs. In CTMSP-Mine, we first propose a transaction clustering algorithm named CO- Smart-CAST to form user clusters based on the mobile transactions using the proposed LBS-Alignment similarity measurement. Then, we utilized the genetic algorithm to generate the most suitable time intervals. To our best knowledge, this is the first work on mining and prediction of mobile behaviors associated with user clusters and temporal relations. A series of experiments were conducted for evaluating the performance of the proposed methods. The experimental results show that CO-Smart-CAST method achieves high-quality clustering results and the proposed CBSS strategy obtains highly precise results for user classification. Meanwhile, our GA-based method obtains the most proper and correct time intervals. For behavior prediction, CTMSP is shown to outperform other prediction methods in terms of precision and F-measure. The experimental results demonstrate that our proposed methods are efficient and accurate under various conditions. In future work, we will apply our method to real data sets. We will also try to integrate CTMSP-Mine with HPM or RMF as a hybrid scheme and design more sophisticated strategies. In addition, we will apply the CTMSP-Mine to other applications, such as GPS navigations, with the aim to enhance precision for predicting user behaviors.

REFERNCES

[1] R. Agrawal, T. Imielinski, and A. Swami, “Mining Association Rule between Sets of Items in Large Databases,” Proc. ACM SIGMOD Conf. Management of Data, pp. 207-216, May 1993.

[2] R. Agrawal and R. Srikant, “Fast Algorithm for Mining Association Rules,” Proc. 20th Int’l Conf. Very Large Databases, pp. 478-499, Sept. 1994.

[3] R. Agrawal and R. Srikant, “Mining Sequential Patterns,” Proc. Int’l Conf. Data Eng., pp. 3-14, Mar. 1995.

[4] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, “Basic Local Alignment Search Tool,” J. Molecular Biology, vol. 215, no. 3, pp. 403-410, Oct. 1990.

[5] A. Ben-Dor and Z. Yakhini, “Clustering Gene Expression Patterns,” J. Computational Biology, vol. 6, no. 3, pp. 281-297, July 1999.

[6] L. Chen and R. Ng, “On the Marriage of Lp-Norms and Edit Distance,” Proc. 30th Int’l Conf. Very Large Databases, pp. 792-803, Aug. 2004.

[7] L. Chen, M. Tamer O ¨ zsu, and V. Oria, “Robust and Fast Similarity Search for Moving Object Trajectories,” Proc. ACM SIGMOD Conf.Management of Data, pp. 491-502, June 2005.

[8] M.-S. Chen, J.-S. Park, and P.S. Yu, “Efficient Data Mining for Path Traversal Patterns,” IEEE Trans. Knowledge and Data Eng., vol. 10, no. 2, pp. 209-221, Apr. 1998.

[9] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with