Prediction of Mobile User Behavior Using Clustering

Sandhya Avasthi, Avinash Dwivedi

Computer Science Department
Krishna Engineering College
Ghaziabad, India
sandhya_avasthi@yahoo.com, avinash_dwivedi@rediffmail.com

Abstract— Services which are recommended to the mobile devices like PDAs, Cellular Phones, smart phones and Laptops while moving using ISAP (Information Service and Application Provider) are increased by accurately predicting user usage pattern. But these discovery may not be always be good enough since the differentiated mobile behaviors among users and temporal periods are not considered simultaneously in the previous works. User relations and temporal property are used simultaneously in this work. Here CTMSP-Mine (Cluster-based Temporal Mobile Sequential Pattern - Mine) algorithm is used to mine CTMSPs. Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) generates user clusters and similarities between mobile sequences are evaluated by Location-Based Service Alignment (LBS-Alignment).The specific time intervals to group the huge mobile logs are found using Genetic Algorithm based method called GetNTSP (Get Number of Time Segmenting Points).CTMSP-Mine technique, which creates CTMSPs utilizes Co-Smart-Cast and time intervals results. These patterns are used to predict the mobile user future behavior and service recommendations are given accordingly.

Index Terms— mining methods and algorithms, mobile environments, Location Based Services, mobile sequential patterns

I. Introduction

Due to the popularity of mobile devices, mobile users can requests services through their mobile devices via Information Service and Application Provider (ISAP) from anywhere anytime [1].The advancement of wireless communication techniques and the popularity of mobile devices such as mobile phones, PDA, and GPS-enabled cellular phones[17], have contributed to a new business model. This business model is known as Mobile Commerce (MC) [2] that provides Location-Based Services (LBS) through mobile phones. MC is expected to be as popular as e-commerce in the future 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 are 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 [2], [3]. Obviously, the behavior pattern, in which the location and the service are inherently coexistent, of mobile users becomes more complex than that of the traditional web systems [4].

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. Fig. 1a shows a moving sequence [2] of a user, where cells are underlined if services are requested there. Fig. 1b shows the record of service transactions, where the service s1 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 prediction of subsequent locations visited by user and user's service requests and service recommendations [5].

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. To achieve a quick response from the system, data mining, which has been used successfully in many applications, is one of the most promising technologies used to fulfill a dynamic service request. Previous works addressed the problem of mining associated service patterns in mobile web network [4]. Previous works also proposed methods to efficiently mine users' sequential mobile access patterns, based on the FP-Tree.

Path traversal patterns for mining mobile web user behaviors have been proposed. 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 it is possible to find the corresponding mobile patterns in each user cluster and time interval.

II. ANALYSIS

A. Existing System

In a mobile network consisting of cells with a base station for each, users of wireless mobile devices move from one location to another in a random manner. The mobile users are served by ISPs and ISAP to access the World Wide Web, to get necessary information in their daily life [17]. When user's movement and their service requests are predicted in advance, it helps to provide customized and efficient service to the users [3]. To help the user get desired information in a short time is one of the promising applications, especially in the mobile environments, where the users do not have much time to surf the web pages. Efficiency is increased to help mobile users experience the usage of web applications and web pages as if they access from a PC. The

Existing system for prediction uses the moving paths of users or the time a user requests for a service. This system does not consider groups of users in mining, but it considers only individual users. This did not provide efficient Prediction of mobile user behavior and it consumes more time to predict and also it lacks in accuracy. Therefore a new system is proposed to solve the problems in prediction.

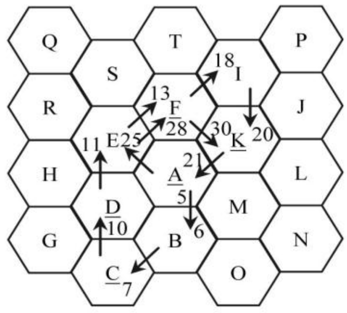
B. Limitations of Existing System

- Most of prediction systems only consider location as prediction parameter.
- The prediction results are inefficient.
- Prediction process consumes more time.
- No precise prediction of mobile user behavior.

C. Proposed System

A novel method, named Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine)[6], for discovering CTMSPs in LBS environment is proposed. In addition, novel prediction strategies are proposed to predict the subsequent user mobile behaviors using the discovered CTMSPs. When mobile users move within the mobile network, the information which includes time, locations, and service requests will be stored in the mobile transaction database. In the data mining mechanism, two techniques and the CTMSP-Mine algorithm are designed to discover the knowledge. First, the CO-Smart-CAST algorithm is proposed to cluster the mobile transaction sequences. In this algorithm, the LBS-Alignment is used to evaluate the similarity [7] of mobile transaction sequences. Second, a Genetic Algorithm based time segmentation method called GetNTSP (Get Number of Time Segmenting Points) [17] to find the most suitable time intervals [8] is introduced. After clustering and segmentation, a user cluster table and a time interval table are generated, respectively. Third, the CTMSP-Mine algorithm is used to mine the CTMSPs from the mobile transaction database according to the user cluster table and the time interval table which are essential in discovering the complete information concerning personal mobile behaviors. The entire procedure of CTMSP-Mine algorithm can be divided into three main steps: 1) Frequent-Transaction Mining, 2) Mobile Transaction Database Transformation and 3) CTMSP Mining.

There are three prediction strategies for selecting the appropriate CTMSP to predict the mobile behaviors of users: 1) the patterns are selected only from the corresponding cluster a user belongs to; 2) the patterns are selected only from the time interval corresponding to current time; and 3) the patterns are selected only from the ones that match the user's recent mobile behaviors. If there exist more than one pattern that satisfies the above conditions, we select the one with the maximal support.



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 Mobile transactions (a) Moving sequences, (b) Mobile Transaction Table.[6]

D. Proposed System Benefits

- It is very precise and efficient
- It consumes less time.
- Considering User clusters and time segmentation simultaneously, complete information concerning personal mobile behaviors is predicted.

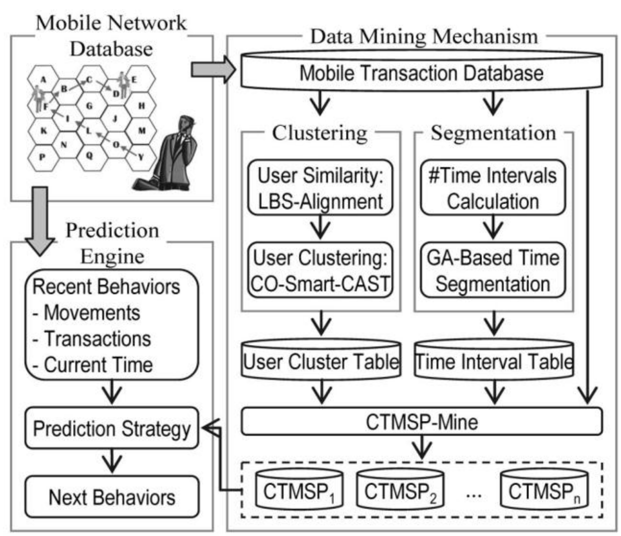


Fig. 2 Overall Proposed System [6]

E. System Features

The system helps to predict the mobile user's behavior. This prediction helps the mobile service providers to enhance their services. The prediction method uses temporal periods and user clusters simultaneously. This methodology provides efficient and precise results.

Such Prediction helps in the following:

i) Related Services can be efficiently recommended [9] to users to help them receive desired information in a short time.

ii) Related items can also be prefetched to reduce the search cost for increasing the access efficiency of Web service servers.

iii) The organization of the services provided by the web can be restructured to fit the mobile user's interest.

III. Mobile Behavior Prediction System

The objective is to predict the mobile user behavior in the near future, to improve the quality of service provided by the Information Service and Application Provider to the wireless mobile device [5].

A. System Implementation

The implementation consists of four modules. The Mobile transactions and behavior generation module logs the complete transactions of mobile users. The clustering and segmentation modules generate the required patterns. The final module uses the new mining technique to predict the future behavior of the mobile users. The modules are described below.

1) Mobile Transactions and Behavior Generation: A mobile network of cells is created with a base station in each cell. Multiple mobile users are created in several cells. The mobile users move in random from one location to another and access more services from the Web through ISAP [1]. The user's movement from one cell to another, their requested services and the current time are monitored and logged into huge databases. Such logs make the Mobile transaction database and it is used in the next two modules.

2) *Usage of Clustering Method:* Users in the different user groups may have different mobile transaction behaviors. Location-Based Service Alignment (LBS-Alignment) technique measures the similarity]among multiple users and provides it to CO-Smart-CAST technique. Clustering method named Cluster Object-based Smart Cluster Affinity [11] to Search Technique (CO-Smart-CAST) builds a cluster model [14].

3) Usage of Segmentation Method: In a mobile transaction database, similar mobile behaviors exist under some certain time segments [8]. To discriminate the different time segments, where the user's behaviors are found to be similar, Get Number of Time Segmenting Points (GetNTSP) algorithm is used. It is a Genetic-Algorithm based method [9].
The segmented time intervals and the corresponding users' mobile transactions are made as entries in the time segmentation table. The output of this module and the second module are provided as input for the fourth module.

4) Applying CTMSP-Mine Technique: User cluster from the second module and time interval table from the third module are used as input in this module. User clustering and temporal property are considered simultaneously, such that the complete mobile sequential patterns are discovered. This module generates patterns called CTMSPs. The patterns and recent behavior of users with current time are provided to the Prediction Strategy. The prediction methods provide the mobile user's behavior in the near future. The Cluster Temporal Mobile Sequential Pattern-Mine algorithm can be divided into three main steps:

- Frequent-Transaction Mining,
- Transaction Database Transformation
- CTMSP Mining.

5) Prediction Technique: The CTMSPs are selected from the corresponding user cluster and time interval.
Three prediction strategies are proposed for selecting the appropriate CTMSP to predict the mobile behaviors of users: 1) the patterns are selected only from the corresponding cluster a user belongs to; 2) the patterns are selected only from the time interval corresponding to current time; and 3) the patterns are selected only from the ones that match the user's recent mobile behaviors.
Whenever a user submits a service request at some place, the user's information including current location, currently requested service and the recent behavior (sequential movement associated with requested services) are input to the prediction component.

B. Algorithm Description

The techniques used for the implementation of the proposed system, like Location-Based Service Alignment (LBS-Alignment), Cluster Object-based Smart Cluster Affinity to Search Technique (CO-Smart-CAST), Get Number of Time Segmenting Points (GetNTSP) algorithm, Cluster Temporal Mobile Sequential Pattern-Mine (CTMSP-Mine) algorithm are described below.

1) LBS Alignment Algorithm: Input data include two mobile transaction sequences [5]. Output data are the similarity between two mobile transaction sequences, with the degrees in the range from 0 to 1. The base similarity score is set as 0.5. Dynamic programming is used to calculate $M_{i,\ j}$ where $M_{i,\ j}$ indicates the value of matrix M in column i and row j, where M is the score matrix of LBS-Alignment [15]. If the locations of two transactions are the same, both the time penalty and the service reward are calculated to measure the similarity score. Otherwise the location penalty is generated to decrease the similarity score. Finally, $M(s.length,s'.length)$ is returned as the similarity score of the two mobile transaction sequences where S and S' are two MTSs. This process goes on till similarity between all MTSs is calculated.

2) The CO-Smart-CAST algorithm: The input data are an N-by-N similarity matrix S. The output data are the clustering result. CO-Smart- CAST can automatically cluster the data according to the similarity matrix without any user-input parameter. First, the CAST method [9] that takes a parameter named affinity threshold t is used as the basic clustering method. Second, use a quality validation method, called Hubert's Γ Statistics, to find the best clustering result. Third, use a hierarchical concept to reduce the sparse clusters. For a clustering result, use Hubert's Γ Statistics to measure its quality by taking the similarity matrix and the clustering result as the input. In each clustering result, calculate its Γ_{obj} and Γ_{clu} which represent the clustering qualities measured by the original object similarity matrix S and the last cluster similarity matrix S_0, respectively. The initial values of S_0 and S are the same since let every object be an independent cluster.
Use the F1 score which is the harmonic mean to combine Γ_{obj} and Γ_{clu} as Γ_{CO}. A higher value of Γ_{CO} represents the better clustering quality. To determine the most suitable t, the easiest way is varying t with a fixed increment and iterating the

executions of CAST to find the best clustering result with the highest Γ_{CO}.
The main drawback of this way is that many iterations of computation are required. For this reason, try to reduce the number of computations by eliminating unnecessary executions, and then, obtain a "near-optimal" clustering result. That is, try to perform a minimal number of CAST executions.

3) The GetNTSP algorithm: The input data are a mobile transaction database D and its time length T. The output data are the number of time segmenting points. For each item, accumulate the total number of occurrences at each time point. Draw a curve of count distribution. Count occurrences of all these time points, 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. In the time point sequence, calculate the average time distance a between two neighboring time points, calculate the number of neighboring time point pairs, in which the time distance is higher than a. The result represents the time segmentation count.

4) CTMSP Mining: User cluster from the second module and time interval table from the third module are used as input in this module. User clustering and temporal property are considered simultaneously, such that the complete mobile sequential patterns are discovered. This module generates patterns called CTMSPs [6]. The patterns and recent behavior of users with current time are provided to the Prediction Strategy. The prediction methods provide the mobile user's behavior in the near future.
There are three prediction strategies for selecting the appropriate CTMSP to predict the mobile behaviors of users: 1) the patterns are selected only from the corresponding cluster a user belongs to; 2) the patterns are selected only from the time interval corresponding to current time; and 3) the patterns are selected only from the ones that match the user's recent mobile behaviors.
The Cluster Temporal Mobile Sequential Pattern-Mine algorithm can be divided into three main steps:

- Frequent-Transaction Mining,
- Mobile Transaction Database Transformation,
- CTMSP Mining.

The Frequent-Transaction mining finds the frequent transactions that are similar. The mobile transaction database is then transformed to remove the noise and outliers. The CTMSP-Mine technique is used to discover the Cluster Temporal Mobile Sequential Patterns. This CTMSP-Mine[6] technique uses the results of CO-Smart-CAST algorithm and GetNTSP algorithm. The CO-Smart-CAST provides the user clusters. Each user cluster is made as an entry in the User cluster table. The GetNTSP algorithm provides the time intervals [8] that have similar mobile user behavior. Each time interval is made as an entry in the Time interval table.
The User cluster table and the Time interval table are provided as input to the CTMSP-Mine Technique. The Cluster Temporal Mobile Sequential Patterns are generated as output of the CTMSP-Mine technique.
The prediction strategy takes the mined patterns CTMSPs and the current transaction of the specific user (like current time, current location, current service) as input. The results of Prediction Strategy provides the future mobile behavior (like location, service) of the specific user. To select the CTMSPs to provide as input to the Prediction Strategy: The user clusters to which the user belongs to, is considered. The CTMSPs that are generated from such user cluster are used as input. The CTMSPs that are generated from the time interval which matches the current time are selected as the input to the Prediction Strategy.

IV. Overview

A. Problem Definition

The problem we are addressing in the proposed system is formulated as follows: Given a user's current mobile transaction sequence S and the current time t_c, the objective is to develop a framework to predict the mobile subsequent behaviors. The aim is to predict the subsequent mobile behaviors using not only S and t_c but also using all the mined CTMSPs.
The problem of CTMSPs mining is formulated as follows: Given a mobile transaction database D containing a large number of mobile transaction sequences of users and a specified support threshold, the problem is to discover all the CTMSPs existing in the database. In this paper, the CTMSP-Mine algorithm and the behavior prediction mechanism are introduced for solving the problem of prediction.

B. Overview of the Project

The services which are provided to the wireless mobile devices (such as PDAs, Cellular Phones, and Laptops) from anywhere, at any time using ISAP (Information Service and Application Provider) are enhanced by mining and prediction of mobile user behaviors. This business model of mobile services is referred as Mobile Commerce [13].
Mining and prediction of mobile movements and associated transactions is the core of the project. The project focuses on discovering mobile patterns from the whole logs. But such discovery may not be precise enough for predictions since the differentiated mobile behaviors among users and temporal periods are not considered simultaneously in the previous works. User relations and temporal property are used simultaneously in this work to provide more accuracy, and scalability.
Prediction strategy is used to predict the subsequent mobile behavior [13]. Here CTMSP-Mine (Cluster-based Temporal Mobile Sequential Pattern - Mine) algorithm is used to mine CTMSPs. In CTMSP-Mine requires user clusters, which are constructed by Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) and similarities [8] between users are evaluated by Location-Based Service Alignment (LBS-Alignment) to construct the user groups.
The temporal property is used by time segmenting the logs using time intervals [7]. The specific time intervals to segment the huge data logs are found using Genetic Algorithm based method called GetNTSP (Get Number of Time Segmenting Points). The

time intervals will have mobile users with similar behavior. For example, the users who access the information about restaurants in the evening. This example have similar users who access the restaurant information. Such users fall in the same time interval. The GetNTSP [7] technique uses the idea behind the genetic algorithm. The user cluster information resulting from CO-Smart-CAST and the time segmentation table are provided as input to CTMSP-Mine technique, which creates CTMSPs. The prediction strategy uses the patterns to predict the mobile user behavior in the near future. The mobile behavior prediction helps ISPs [1] to improve their QoS given to mobile users.

VI. Simulation

A. Simulation Tool for Prediction

The Data Mining Simulation tool Matlab 10 is used. This simulation tool could be used in multiple platforms. This tool could utilize data accesses such as ODBC, Excel, plain ASCII files. The data preprocessing methods used here are pick mix, sampling, partitioning, field reordering, table fusion, recoding numeric intervals. The association rules are mined using Apriori [16] method. The visualization methods are statistical graphics, 3D, animation, interactive and hierarchical navigation.

VII. Conclusion

In this paper, a new method is proposed and named as CTMSP-Mine, for discovering CTMSPs in LBS environments. Further, prediction strategies to predict the subsequent user mobile behaviors using the discovered CTMSPs are introduced. In CTMSP-Mine technique, transaction clustering algorithm named CO Smart-CAST is used to form user clusters, based on the mobile transactions using the proposed LBS-Alignment similarity measurement. Then, GetNTSP method is utilized to generate the most suitable time intervals. Using temporal periods and user clusters simultaneously for prediction, enhances the prediction results. Such prediction results are used by the corresponding mobile service providers to enhance their services.

V. Future Work

By implementing prioritization, it is possible to provide priorities for selected users among the complex user behavior. Many users utilize the mobile services every day but their interest and priorities are different from other user. Such users are prioritized over other mobile users. These prioritized services help to satisfy the needs of mobile users completely when resources are limited. Further, this could be implemented for group behavior too.

References

[1] J. Veijalainen, “Transaction in Mobile Electronic Commerce,” Proc. Eighth Int’l Workshop Foundations of Models and Languages for Data and Objects, pp. 203-227, Sept. 1999.

[2] U. Varshney, R.J. Vetter, and R. Kalakota, “Mobile Commerce: A New Frontier,” Computer, vol. 33, no. 10, pp. 32-38, Oct. 2000.

[3] Chen, T.S., Chou, Y.S., Chen, T.-C, (2011), “*Mining User Movement* Behavior Patterns in a Mobile Service Environment”, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE.

[4] Tseng, V.S.M.; Lin, K.W.C, (2005) “*Mining sequential mobile access patterns efficiently in mobile web systems*”, Advanced Information Networking and Applications, AINA.

[5] J.B. Schafer, J. Konstan, and J. Riedl, “Recommender Systems in E-Commerce,” Proc. ACM Conf. Electronic Commerce, pp. 158-166, Nov. 1999.

[6] Lu, E.H., Tseng, V.S., Yu, P.S. (2011), “*Mining Cluster-Based temporal* Mobile Sequential Patterns in Location-Based Service Environments”, Knowledge and Data Engineering, IEEE Transactions.

[7] 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.

[8] M. Halvey, T. Keane, and B. Smyth, “Time-Based Segmentation of Log Data for User Navigation Prediction in Personalization,” Proc. IEEE Int’l Conf. Web Intelligence, pp. 636-640, Sept. 2005.

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

[10] Baoyao Zhou, Siu Cheung Hui, Kuiyu Chang, (2006), “*Enhancing* mobile Web access using intelligent recommendations”, Intelligent Systems, IEEE.

[11] V.S. Tseng and C. Kao, “Efficiently Mining Gene Expression Data via a Novel Parameter less Clustering Method,” IEEE/ACM Trans. Computational Biology and Bioinformatics, vol. 2, no. 4, pp. 355-365, Oct.-Dec. 2005.

[12] Tseng, V.S., Lu, R.H.-C., Cheng-Hsien Huang (2007), “*Mining temporal mobile sequential patterns in location-based service environments*”, Parallel and Distributed Systems.

[13] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, “WhereNext: A Location Predictor on Trajectory Pattern Mining,” Proc. 15thInt’l Conf. Knowledge Discovery and Data Mining, pp. 637-646, June 2009

[14] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” Proc. Second Int’l Conf. Knowledge Discovery and Data Mining, pp. 226-231, Aug. 1996.

[15] 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.

[16] R. Agrawal and R. Srikant, “Mining Sequential Patterns,” *Proc. 11th Int’l Conf. Data Eng.*, IEEE CS Press, 1995, pp. 3–14.

[17] Balamurugan.V”Mining User Mobile Behavier in Location Based Services” IJ SRP, Volume 2, Issue 9, September 2012 1 ISSN 2250-3153