A Mining Cluster Based Temporal Mobile Sequential Patterns in Location Based Service Environments using CTMSP Mining

Shamshedbanu.M[1] Navin Kumar.R[2]
[2]Assistant Professor
[1,2]Department of Master of Computer Application
[1,2]Nandha Engineering College,Erode,Tamilnadu

Abstract— Due to a wide range of potential applications, research on mobile commerce has received a lot of interests from both of the industry and academia. Among them, one of the active topic areas is the mining and prediction of users' mobile commerce behaviors such as their movements and purchase transactions. The project proposes a novel framework, called Mobile Commerce Explorer (MCE), for mining and prediction of mobile users' movements and purchase transactions under the context of mobile commerce. The proposed framework is termed as MCE (Mobile Commerce Explorer) which consists of three major components: 1) Similarity Inference Model (SIM) for measuring the similarities among stores and items, which are two basic mobile commerce entities considered in this paper; 2) Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm for efficient discovery of mobile users' Personal Mobile Commerce Patterns (PMCPs); and 3) Mobile Commerce Behavior Predictor (MCBP) for prediction of possible mobile user behaviors. In addition, 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.

Key words: MCE, LBS, PMCP, CTMSP Mining

I. INTRODUCTION

LOCATION-BASED services (LBS) have attracted significant attention from both industrial and academic Communities. Many LBS services such as Foursquare (foursquare.com) and Google Maps (maps.google.com) have been widely accepted because they can provide users with location-aware experiences. Existing LBS systems employ a pull model or user-initiated model where a user issues a query to a server which responds with location aware answers. For example, if a mobile user wants to find seafood restaurants nearby, she issues a query "seafood restaurant" to an LBS system, which returns answers based on user's location and keywords.

To provide users with instant replies, a push model or server-initiated model is becoming an inevitable computing model in next-generation location-based services. In the push model, subscribers register spatial-textual subscriptions to capture their interests, and publishers post spatiotextual messages. These calls for a high-performance location-aware publish/subscribe system to deliver messages to relevant subscribers. This computing model brings new user experiences to mobile users, and can help users retrieve information without explicitly issuing a query.

There are many real-world applications using location aware publish/subscribe services. The first one is Group on. Group on customers register their interests with locations and keywords. For each Group on message (e.g., "iphone4s AT&T package" at Manhattan), the system provider sends the message to the customers who may be potentially interested in the message by evaluating the spatial proximity and textual relevancy between subscriptions and the message. The second one is location-aware AdSense, which extends traditional AdSense to support location-aware services. The advertisers register their location-based advertisements (e.g., "seafood" at Manhattan) in the system.

II. EXISTING SYSTEM

Clustering mobile transaction data helps in the discovery of social groups, which are used in applications such as targeted advertising, shared data allocation, and personalization of content services. In previous studies, users are typically clustered according to their personal profiles (e.g., age, gender, 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, the evaluation of similarities of mobile transaction sequences (MTSs) is required.

The existing system consists of a novel framework, called Mobile Commerce Explorer (MCE), for mining and prediction of mobile users' movements under the context of mobile commerce. The MCE framework consists of three major components:

1) Similarity Inference Model (SIM) for measuring the similarities among stores and items, which are two basic mobile commerce entities considered in this paper;
2) Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm for efficient discovery of mobile users' Personal Mobile Commerce Patterns (PMCPs); and
3) Mobile Commerce Behavior Predictor (MCBP) for prediction of possible mobile user behaviors.

A. *Demerits:*

The existing system has following disadvantages,

- Discovering mobile patterns from the whole logs are not precise enough for predictions.
- Differentiated mobile behaviors among users and temporal periods are not considered.
- Request the users to set up some parameters which is difficult to determine.
- Not applicable in the LBS (Location Based Service) scenario.
- Process data with spatial similarity measures, while clustering methods with non-spatial similarity measures are required for LBS environments.

III. PROPOSED SYSTEM

The proposed system develops a novel algorithm, namely, Cluster-based Temporal Mobile Sequential Pattern Mine (CTMSP-Mine), to discover the Cluster-based Temporal

Mobile Sequential Patterns (CTMSPs). Since, a prediction strategy is proposed to predict the subsequent mobile behaviors, in CTMSP-Mine, user clusters are constructed by a novel algorithm named Cluster Affinity Search Technique (CAST) and similarities between users are evaluated by the proposed measure, Location-Based Service Alignment (LBS-Alignment). At the same time, a time segmentation approach is presented to find segmenting time intervals where similar mobile characteristics exist. The project considers mining and prediction of mobile behaviors with considerations of user relations and temporal property simultaneously.

A. *Merits:*

The proposed system has following advantages,

- Predicts the subsequent user mobile behaviors effectively.
- Generate the most suitable time intervals for time segmentation.
- Mines and predicts the mobile behaviors with considerations of user relations and temporal property simultaneously.
- Suitable for Location-Based Service Environments.

IV. PROJECT DESCRIPTION

A. *Problem Definition:*

Clustering mobile transaction data helps in the discovery of social groups, which are used in applications such as targeted advertising, shared data allocation, and personalization of content services. In previous studies, users are typically clustered according to their personal profiles (e.g., age, gender, 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, the evaluation of similarities of mobile transaction sequences (MTSs) is required.

The existing system consists of a novel framework, called Mobile Commerce Explorer (MCE), for mining and prediction of mobile users' movements under the context of mobile commerce. The MCE framework consists of three major components:

1) Similarity Inference Model (SIM) for measuring the similarities among stores and items, which are two basic mobile commerce entities considered in this paper;
2) Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm for efficient discovery of mobile users' Personal Mobile Commerce Patterns (PMCPs); and
3) Mobile Commerce Behavior Predictor (MCBP) for prediction of possible mobile user behaviors.

B. *Overview of The Project:*

Due to a wide range of potential applications, research on mobile commerce has received a lot of interests from both of the industry and academia. Among them, one of the active topic areas is the mining and prediction of users' mobile commerce behaviors such as their movements and purchase transactions. The project proposes a novel framework, called Mobile Commerce Explorer (MCE), for mining and prediction of mobile users' movements and purchase transactions under the context of mobile commerce.

The proposed framework is termed as MCE (Mobile Commerce Explorer) which consists of three major components: 1) Similarity Inference Model (SIM) for measuring the similarities among stores and items, which are two basic mobile commerce entities considered in this paper; 2) Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm for efficient discovery of mobile users' Personal Mobile Commerce Patterns (PMCPs); and 3) Mobile Commerce Behavior Predictor (MCBP) for prediction of possible mobile user behaviors.In addition, 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. The project proposes 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 Affinity Search Technique (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. To the best knowledge, this is the first work on mining and prediction of mobile behaviors with considerations of user relations and temporal property simultaneously. Through experimental evaluation under various simulated conditions, the proposed methods are shown to deliver excellent performance.

V. PROJECT IMPLEMENTATION

The project contains the following modules

1) Item Similarity
2) Store Similarity
3) Clustering of mobile transaction sequences.
4) Time segmentation of mobile transaction sequences.
5) Discovery of CTMSPs.

A. *Item Similarity:*

In this module, the items are added with their score values. Then similarity between two items are found out using the given similarity formula. Likewise similarity matrix is found out for all the items in the list.

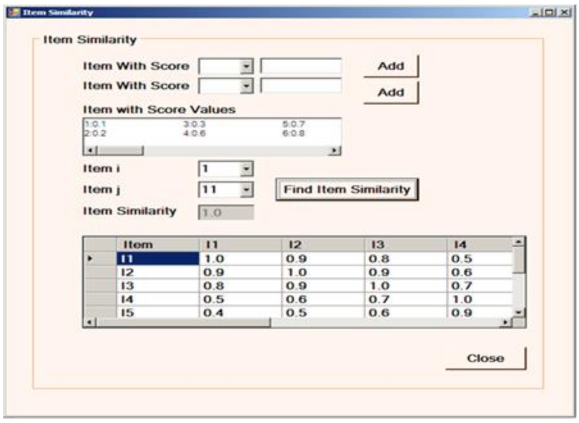

Fig. 1: Item Similarity Matrix

B. Store Similarity:

In this module, the stores are added with their score values. Then similarity between two store are found out using the given similarity formula. Likewise similarity matrix is found out for all the stores in the list.

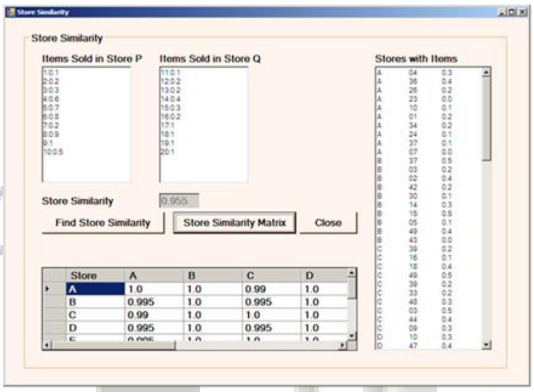

Fig. 2: Store Similarity Matrix

C. Clustering Of Mobile Transaction Sequences:

In a mobile transaction database, users in the different user groups may have different mobile transaction behaviors. The first task to tackle is to cluster mobile transaction sequences. In this module, a parameter-less clustering algorithm called CAST is proposed. Before performing the CAST, a similarity matrix S is to be generated, 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 to tackle is to measure the content similarity between mobile transactions.

The LBS-Alignment algorithm is proposed, which can obtain the similarity. LBS-Alignment is based on the consideration that two mobile transaction sequences are more similar, when the orders and timestamps of their mobile transactions are more similar. CAST algorithm is used to cluster the users.

D. 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.

A new time segmentation method is proposed to automatically obtain the most suitable time segmentation table with common mobile behaviors. The algorithm below shows the procedure of the proposed time segmentation method, named Get Number of Time Segmenting Points (GetNTSP) algorithm.

E. Discovery Of Ctmsp:

In order to mine the cluster-based temporal mobile sequential patterns efficiently, we proposed a novel method named CTMSP-Mine to achieve this mining procedure. In CTMSP-Mine, both factors of user cluster and time interval are taken into account such that the complete mobile sequential patterns can be discovered. The entire procedures of CTMSP-Mine algorithm can be divided into three main steps:

1) Frequent-Transaction Mining,
2) Mobile Transaction Database Transformation, and
3) CTMSP Mining.

1) Frequent-Transaction Mining:

In this phase, the frequent transactions (F-Transactions) are mined in each user cluster and time interval by applying a modified Prior algorithm.

2) Mobile Transaction Database Transformation:

In this phase, F-Transactions are used to transform each mobile transaction sequence S into a frequent mobile transaction sequence S'. According to Table 3, if a transaction T in S is frequent, T would be transformed into the corresponding F-Transaction.

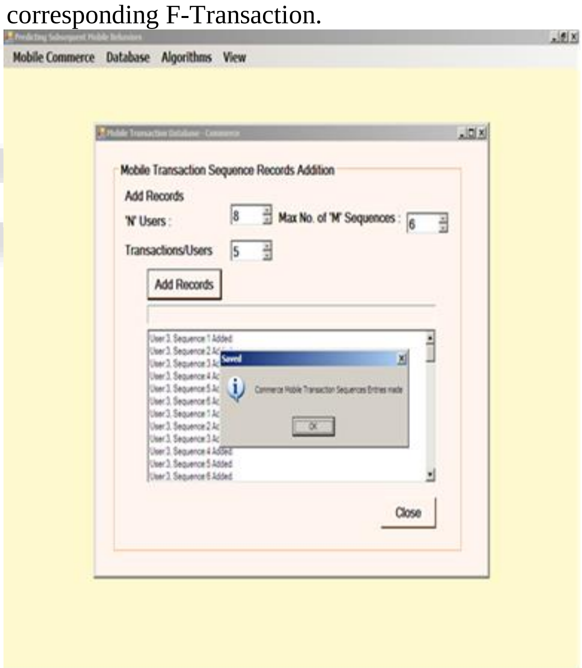

Fig. 3: Mobile Transaction Database

3) Ctmsp Mining:

In this phase, all the CTMSPs are mined from the frequent mobile transaction database. Frequent 1-CTMSPs are obtained in the frequent-transaction mining phase. In the mining algorithm, we utilize a two-level tree named Cluster-based Temporal Mobile Sequential Pattern Tree (CTMSP-Tree). The internal nodes in the tree store the frequent mobile transactions, and the leaf nodes store the corresponding paths. Moreover, every parent node of a leaf node is designed as a hash table which stores the combinations of user cluster tables and time interval tables.

VI. DATABASE DESIGN

The most important consideration in designing the database is how information will be used. The main objectives of designing a database are:

A. Data Integration:

In a database, information from several files are coordinated, accessed and operated upon as through it is in a single file. Logically, the information are centralized, physically, the data may be located on different devices, connected through data communication facilities.

B. Data Integrity:

Data integrity means storing all data in one place only and how each application to access it. This approach results in more consistent information, one update being sufficient to achieve a new record status for all applications, which use it. This leads to less data redundancy; data items need not be duplicated; a reduction in the direct access storage requirement.

C. Data Independence:

Data independence is the insulation of application programs from changing aspects of physical data organization. This objective seeks to allow changes in the content and organization of physical data without reprogramming of applications and to allow modifications to application programs without reorganizing the physical data.

The tables needed for each module were designed and the specification of each and every column was given based on the records and details collected during record specification of the system study

D. Table: Admin:

FIELDNAME	TYPE	SIZE	DESCRIPTION
USERNAME	VARCHAR	15	Primary Key
PASSWORD	VARCHAR	15	

TABLE 1: Purpose: To store the username and passwords of the application users.

E. Table: MT sequence:

FIELDNAME	TYPE	SIZE	DESCRIPTION
SNO	INT	4	Primary Key
USERID	SMALLINT	2	
TIMEINDEX	SMALLINT	2	
LOCATIONID	VARCHAR	2	
SERVICEID	VARCHAR	50	

TABLE 2: Purpose: To Store the Mobile Transaction Sequence Records.

VII. CONCLUSION

A novel method named CTMSP-Mine is proposed, for discovering CTMSPs in LBS environments. Furthermore, novel prediction strategies are proposed to predict the subsequent user mobile behaviors using the discovered CTMSPs. In CTMSP-Mine, first a transaction clustering algorithm is proposed named CO-Smart-CAST to form user clusters based on the mobile transactions using the proposed LBS-Alignment similarity measurement. Then, the time segmentation algorithm is utilized to generate the most suitable time intervals. To our best 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, the algorithms obtain 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 the proposed methods are efficient and accurate under various conditions. The application works well for given tasks in windows environment. Any node with .Net framework installed can execute the application and identifies the best site. The underlying mechanism can be extended to any / all kind of web servers and even in multi-platform like Linux, Solaris and more. The system is planned to extend the services can be given as input to IBM architecture also. The system eliminates the difficulties in the existing system. It is developed in a user-friendly manner. The system is very fast in applying algorithm. This software is very particular in predict the subsequent mobile behaviors.

VIII. SCOPE FOR FUTURE DEVELOPMENT

The new system become useful if the below enhancements are made in future. In future work, the method can be applied to real data sets. In addition, the CTMSP-Mine can be applied to other applications, such as GPS navigations, with the aim to enhance precision for predicting user behaviors. The application if developed as web site can be used from anywhere. The new system is designed such that those enhancements can be integrated with current modules easily with less integration work.

REFERENCES

[1] Butterworth R., Piatetsky-Shapiro G. and Simovici D.A., ‘On Feature Selection through Clustering’, In Proceedings of the Fifth IEEE international Conference on Data Mining, pp 581-584, 2005.
[2] Fleuret. F (2004), ‘Fast Binary Feature Selection with Conditional Mutual Information’, Journal of Machine Learning Research 5 (2004) 1531–1555.
[3] Huan Liu and Lei Yu, ‘Towards Integrating Feature Selection Algorithms for Classification and Clustering’, IEEE Transactions on Knowledge and Data Engineering Vol. 17 No. 4, 2005.
[4] Huan Liu, Manoranjan Dash and Hiroshi Motoda, ‘Feature Selection Using Consistency Measure’, Second International Conference, DS’99 Tokyo, Japan, December 6–8, 1999 Proceedings, pp 319-320.
[5] Huan Liu and Rudy Setiono, ‘A Probabilistic approach to Feature Selection – A Filter approach’, 2000.
[6] Huan Liu, Hiroshi Motoda (2002), and Lei Yu, ‘Feature Selection with Selective Sampling’, Proceedings of 19th International Conference on Machine Learning, pp. 395-402, 2002.
[7] Huan Liu and Lei Yu (2003), ‘Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution’, Proceedings of 20th International Conference on Machine Learning, pp. 856-863, 2003.
[8] Huan Liu and Lei Yu (2005), ‘Toward Integrating Feature Selection Algorithms for Classification and

Clustering', IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 4, April 2005.
[9] I. Kononenko (1994), 'Estimating Attributes: Analysis and Extension of RELIEF', Proceedings of Sixth European Conference on Machine Learning, pp. 171-182, 1994.
[10] K. Kira and L.A. Rendell (1992), 'The Feature Selection Problem: Traditional Methods and a New Algorithm', Proceedings of 10th National Conference on Artificial Intelligence, pp. 129-134, 1992.
[11] M.A. Hall (2000), 'Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning', Proceedings of 17th International Conference on Machine Learning, pp. 359-366, 2000.
[12] Liu H., Motoda H. and Yu L., 'Selective sampling approach to active feature selection', Artif. Intell., 159(1-2), pp 49-74 (2004).
[13] Park H. and Kwon H., 'Extended Relief Algorithms in Instance-Based Feature Filtering', In Proceedings of the Sixth International Conference on Advanced Language Processing and Web Information Technology, pp 123-128, 2007.
[14] Pereira F., Tishby N. and Lee L., 'Distributional clustering of English words, In Proceedings of the 31st Annual Meeting on Association For Computational Linguistics', pp 183-190, 1993.
[15] Qinbao Song, Jingjie Ni and Guangtao Wang (2013), 'A Fast Clustering-Based Feature Subset Selection Algorithm for High Dimensional Data', IEEE Transactions on Knowledge and Data Engineering Vol: 25 No: 1 Year 2013.
[16] Sha. C, Qiu X. and Zhou A (2008), 'Feature Selection Based on a New Dependency Measure', Fifth International Conference on Fuzzy.
[17] Sunita Beniwal and Jitendar Arora (2012), 'Classification and Feature Selection Techniques in Data Mining', International Journal of Engineering Research and Technology, Volume 1 Issue 6, August 2012.
[18] Yuxuan SUN, Xiaojun LOU and Bisai BAO., 'A Novel Relief Feature Selection Algorithm based on Mean-Variance Model', Journal of Information and Computational Science, pp 3921-3929, 2011.
[19] Zhao Z. And Liu H. (2009), 'Searching for Interacting Features in Subset Selection', Journal of Intelligent Data Analysis, 13(2), pp 207-228, 2009.
[20] Zhao Z. and Liu H., Searching for interacting features, In Proceedings of the 20th International Joint Conference on AI, 2007.